DOCUMENT RESUME

ED 051 250 24 TE 499 827

AUTHOR Kensler, Gordon
TITLE Preconference Education Research Training Program for Descriptive Research in Art Education. Final Report.
INSTITUTION National Art Education Association, Washington, D.C.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-0-0071
PUB DATE Jan 71
GRANT OEG-0-70-2849
NOTE 109p.

EDRS PRICE EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS *Art Education, Behavioral Objectives, *Change Agents, *Data Collection, Institutes (Training Programs), *Research Methodology, *Teacher Attitudes
IDENTIFIERS National Art Education Association

ABSTRACT

The purpose of this publication is to inspire an interest in research and encourage art teachers at all levels to seek more informaticn and try this methodology (descriptive research in art education) in their own teaching environment. There are some who see art and science as quite distinctly different approaches to reality. Another view is that art and science come together at the point of active discovery. This presentation is made with two observations in mind: (1) Many art educators distrust science generally; (2) Many tend to be wholistic in personality. The document is divided into five sections. It is concluded that art educators trained in research methods are relatively few and their combined number cannot provide the descriptive data that is needed. It is hoped that more interest and active involvement in gathering descriptive material will come from teachers and supervisors working in the schools and colleges. Five appendices are attached. (CK)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR 0-0071
PA 24
TI/0

ED051250

FINAL REPORT
Project No. 0-0017 71
Grant No. OEG-0-70-2849

PRECONFERENCE EDUCATION RESEARCH TRAINING PROGRAM FOR DESCRIPTIVE RESEARCH IN ART EDUCATION

Gordon Kensler
National Art Education Association
1201 - 16th Street, N.W.
Washington, D.C. 20036

TE 499 827

January, 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

PREFACE i

INTRODUCTION 1

1. OBSERVATION: THE BASIS FOR TEACHING AND RESEARCH 6
Dale B. Harris

Can We Apply Science to Art 6
Observation: The Basis of Science 7
Characteristics of Science and Art 8
The Identification of Criteria 9
Classification: The Application of Criteria 11
The Identification of Variables 11
The Analytic Method 13
An Example of a Quantatative Scale 13
An Example of a Qualitative Scale 16
Summary 19

11. VARIETIES OF OBSERVATION 20
Dale B. Harris

The Idea of Dimension 21
Identifying Meaningful Dimensions 23
The General Observations 23
Observing the Stream of Behavior 24
Making Judgments in General Observation 26
The Selective Observation 26
Tests and Work Samples 28

111. THE CASE STUDY 29
Dale B. Harris and Kenneth R. Beittel

A. The Case Method in Art Education (D. B. H.) 29
The Case Method Gives Particular Information 29
A Study of Drawing Process 30
Other Uses of the Case Method 30
Studies of Growth in Interest of Skill 31
An Example of a Longitudinal Study 31

B. Some Value of Case Studies (K. R. B.) 50

IV. IDENTIFYING A RESEARCHABLE PROBLEM 53
Dale B. Harris

Formulating a Problem Statement 53
A Sample Question 53
Restating the Initial Question 54
Implicit Assumptions 54
Formulating Specific Substudies 54
A Checklist of Observable Behaviors 55
A Second Example 56
Determining Criteria 57
Reliability 58
A Possible Experiment 58
Correlation Studies 59

V. AN INTRODUCTION TO MEASUREMENT 60
William Rabinowitz

Fundamental Concepts of Measurement 60
Nominal Scales 62
Ordinal Scales 63
Interval Scales 64
Ratio Scales 65
Errors in Measurement 66
Systematic Errors 67
Random Error 70

POSTSCRIPT 73

BIBLIOGRAPHY 76

APPENDIX 78

PREFACE

This report is one outcome of the Preconference Educational Research Training Institutes that were offered by the National Art Education Association with the support and encouragement of the U.S. Office of Education.

As a beginning it may be helpful to establish a sense of how the Institutes were organized, with a brief resume of the development of events that took place. Those who are interested in more detailed information will find it in the appendices as indicated. The introduction in this report provides a background of information for the content that is presented thereafter.

In the fall of 1969 an announcement of the institutes and a call for participants was issued to all NAEA members through the NAEA Newsletter, ART EDUCATION, and the NEA Reporter. For those outside the membership, direct mail announcements were sent to college and university art departments and state departments of education. Since the number of applications did not exceed the number that could be accommodated, no selection was necessary, and all who applied became participants. Each Institute had 25-30 participants composed of elementary classroom teachers, art teachers, consultants, and supervisors from the public schools and art education at all levels of Higher Education. Although the training activity was designed to benefit art educators who had limited or no background in research, a small portion of the participants had advanced degrees and used the program to supplement their skills in research in an area they felt an important need.

The Institutes were held during the three days immediately prior to the regular Regional Conventions of the National Art Education Association in 1970. Approximately eight hours of each day were scheduled for instruction and working on problems*. Large group sessions were designated primarily for instruction, but open question and answer periods and reports from the small groups also took place during this time. The participants were randomly placed into four or five small groups for work sessions. The work sessions started with assigned problems and later became a time when participants worked on problems of their own choosing. Light assignments for "homework" were also given as preparation for the activities of the following day.

* Appendix B

The training activity concentrated on providing the participants with a workable concept of descriptive research and the fundamental techniques of observation methodology that could be used in their respective situations. As a result of this training, the participants acquired skill in (1) problem definition and descriptive research, (2) selecting research design for descriptive studies, and (3) the methodology for implementing descriptive study using observation techniques.

The importance of these Institutes and their influence in the field of art education is difficult to assess. The evaluator for the Institutes had a dual responsibility. He provided formative evaluation by maintaining ongoing communication with the participants as to how they were responding to the instruction and then feeding this information back to the instructional staff each day. He administered pre and post test* on the content of the Institutes and a final subjective Institute evaluation that provided the basis for summative evaluation**. However, the true measure of the success of the Institutes will come from the increased knowledge contributed to the field from projects and influential programs initiated or enhanced by participation in the Institutes. Basic practical and theoretical knowledge used or transmitted as a result of participation can provide a valuable and confirming input into the field.

The main part of this report is in essence a transcription of the presentations that were given by the primary instructional staff at the Institutes. Some of the material that was accompanied by slides and the short presentations that were mainly for instructing and participants have been omitted. Unfortunately, the "give and take" of shared dialogue and professional interaction, the laughs and the arguments that occurred especially in the small groups which brought a sense of camaraderie to the Institutes, lose all flavor in transcription. What appears here is a consolidation of material that was presented in shorter segments and has been rearranged into related sections for more comprehensive reading.

No attempt has been made to edit out the individual styles of the contributors in order to make the text more uniform. Since the manuscripts were prepared mainly from tape recordings of the presentations, a conversational quality is often apparent. This quality, combined with the color and texture of the individual styles, provides a fresh approach to subject matter that is usually couched in esoteric language.

* Appendicies D and E
** Appendix C

While this publication may be considered complete in itself, it may also serve as an introduction to one of the methods for research in art education. For the beginner we hope this may open the door to new understandings and activity, and for those trained in research it may serve as a companion and desk reference in their professional undertakings.

The planning and presentation of the Institutes and the preparation of this publication have required the time and service of the many people listed below, and these contributions are most appreciated. A special acknowledgement should be made to Dr. Dale B. Harris and Dr. William Rabinowitz who so ably related their areas of expertise to the problems of art education. In addition they graciously contributed many hours to the consolidation and editing of the transcriptions of their presentations in order to make this publication possible. Dr. Kenneth Beittel made presentations at two of the Institutes with slides on his current work on drawing using case study technique. His section was especially written for this publication.

INSTITUTE PLANNING COMMITTEE

Dr. Harlan Hoffa, Chairman, Department of Art Education, Indiana University.
Dr. Gordon L. Kensler, Project Director, Associate Professor of Art Education and Research Associate for the Institute for Community Art Studies, University of Oregon.
Dr. Stanley Madeja, Director, Aesthetic Education Project, Central Midwestern Regional Educational Laboratory, Inc.
Dr. Edward Mattil, Head, Department of Art Education, Pennsylvania State University.

NAEA STAFF

Dr. Charles M. Dorn, Executive Secretary, National Art Education Association.
Dr. John Mahlmann, NAEA Project Coordinator, Assistant Executive Secretary, National Art Education Association.

INSTITUTE STAFF

Dr. Gordon L. Kensler, Project Director, Associate Professor of Art Education and Research Associate for the Institute for Community Art Studies, University of Oregon.
Dr. Dale Harris, Primary Instructor and Research Specialist, Professor, Department of Psychology, Pennsylvania State University.

Dr. William Rabinowitz, Evaluator and Instructor, Head, Department of Educational Psychology, Pennsylvania State University.

Dr. Kenneth R. Beittel, Instructor, Professor of Art Education, Pennsylvania State University.

Dr. Marylou Kuhn, Instructor, Professor of Art Education, Florida State University.

Dr. David Templeton, Instructor, Associate Professor, Department of Art, Western Washington State College.

Dr. Brent Wilson, Instructor, Associate Professor of Art and Education, University of Iowa.

INTRODUCTION

The National Art Education Association has a growing reputation for providing a means of implementing change and innovation in the field. In 1967 the national office invited a representative from each of the four NAEA regionals along with several other well known art educators to attend a planning meeting for the purpose of organizing some training institutes for the membership. In 1968 the NAEA in cooperation with the Bureau of Research in the U. S. Office of Education, offered the first Research Training Institutes in the history of the Association. These institutes were offered immediately prior to the regional conventions and focused on the development of behavioral objectives in art for use in research and curriculum construction. Their enthusiastic acceptance by the participants prompted the planning committee to provide for a continuance of this training with another institute prior to the national conference in 1969.

Late in the spring of 1969, the planning group met to assess the results of the two previous training institutes on behavioral objectives and to ascertain desirable directions for future programs. Two decisions were made in the early part of the two-day meeting. The committee was highly supportive of the work that had been done on behavioral objectives, and the reports and evaluations from these institutes offered credence to the notion that training institutes were indeed a useful technique for providing a direct input of ideas for effective change. The decision to continue offering training institutes was unanimous. The second decision required some deliberation, but because of the unlikelihood that more than one kind of institute would be funded, the committee decided that the next one should focus on another topic in order to reach a different population within the membership.

The decision to use the topic of descriptive research for the recent training institutes came after careful consideration of the needs of the field of art education and the establishment of some priorities. The outcome was two proposals for different kinds of training institutes that would serve different kinds of members. Unfortunately, only the one on descriptive research received funding from the U. S. Office of Education.

The purpose of these Institutes is easily inferred from the title, and one would expect that some kind of training related to descriptive research was presented. What may remain obscure to some is the rationale for having this kind of training; why is it important, and what

relevance does it have for the field of art education? The framework of information that follows is presented with the notion that without understanding the need for descriptive research the importance of the content of this institute is not fully recognized.

Art education in recent years has experienced an intensified interest in the area of research. The leadership role in the organization and dissemination of research information have been assumed by the National Art Education Association. With the creation of the journal, Studies in Art Education, first published in 1959, a systemized method of reporting research to the field was established.

The monthly journal Art Education periodically deals with research topics and is distributed to every member of the organization. Recently the Lowenfeld Series of Research Monograph has been published through the national organization, to deal specifically with research information. Other research studies and information come from individual institutions and are less widely distributed.

One of the current problems in the field is the minimal interest and irregular implementation of research findings by supervisors and classroom teachers. Attention to the problem is noted in the introduction to the recent publication, Exemplary Programs in Art Education:

> The gap between the theoretician and the art teacher became apparent as the programs were reviewed. Many art teachers and art administrators feel that theory and research have very little, if any, relationship to the teaching of art within the classroom. The skepticism of the techniques of the theoretician may be attributed to a lack of understanding of such procedures by the majority of art specialists and also the lack of dialogue between theorist and art teachers. (10)

In the art area specifically, those working in the classrooms must generate research ideas which are relevant to their needs as teachers if pertinent problems are to be identified for further study and the overall improvement of instruction.

The body of knowledge concerned with research in the visual arts has been reviewed by Hastie in the Yearbook of the National Association for the Study on Art Education,(4) and by Hausman in the Report of the Commission on Art Education, (17) and it is systematically reported in bulletins and journals relating directly to the visual arts and education. A number of concentrated efforts on the national level have been made to assimilate and correlate this existing research into a meaningful body of knowledge, such as, the Seminar in Art Education for Research and Curriculum Development,(1) under the direction of Mattil, and the

Seminar on Elementary and Secondary School Education in the Visual Arts,(20) directed by Conant. However, the existence of this knowledge and the attempts at correlation to classroom practice are seemingly not apparent to the art teacher. The lack of awareness of this progress in the research area may be a contributing factor to the lack of enthusiasm by the art teacher towards research in his field.

The major problem seems to be that neither the art specialist nor the classroom teacher has developed the skills for (1) translating research findings into classroom practice, or (2) initiating and carrying out research projects that will provide useful information for improving their programs.

In addition, many institutions of higher education are not providing for the development of these skills in their programs of teacher training.

In viewing the present status of art education research in the public schools, it is important to recognize that the climate for conducting research within the context of the public school is not a positive one. Lack of time for research activity, lack of available funds, lack of knowledge of research procedures and implications, and lack of interest on the part of school boards, administrators, and teachers have resulted in a negligable amount of research conducted by the classroom teacher.

Most researchers would agree that if these conditions exist they would hinder the implementation of any type of research activity. However, these deficiencies will cease to exist only when the subject area specialist initiates some enthusiasm for research at the classroom level. Specifically in the art area, the art specialist in the public school and the junior college must initiate research ideas which would be relevant to their problems within the classroom situations. There has been no coherent outcry for help from art teachers to study problems which are significant to the instruction of art. In fact, the converse might be nearer the true picture. That is, funds and opportunities are available to the classroom teacher for conducting research in the visual arts, but few are taking advantage of current programs to initiate plans of study.

If art education as a discipline is to conduct a total research program which is meaningful to all the individuals in the field, it must begin to integrate the classroom teacher into some phase of the program. Most research today is initiated and conducted in art education by persons directly or indirectly working with some institution of higher learning. It might be a better rounded program gaining more popular support if some part of the research was initiated solely within the context of the school and specifically designed to investigate a problem which has relevance to the teacher and the instructional program.

Most educational researchers accept the fact that the purpose of research in education is a means of expanding subject area and ultimately improving instruction by application of new found knowledge. The role which the researcher has assumed to be important is that without various types of research activity to indicate new directions, art education could easily become a static discipline. It is evident that there are some deficiencies at present in the relationships between the researcher and the art teacher as to how the work of one offers potentially useful contributions to the work of the other.

Research in art education, as in other fields, encompasses a range of inquiries with a variety of purposes and procedures that may be categorized under the titles of analytical, experimental, and descriptive. Analytic research includes historical and philosophical analyses as well as other primarily deductive systems that can be used to derive relationships that are not necessarily of an empirical nature. Experimental Research is devoted to the study of "casual" relationships, and is often used in evaluating the advantages of one teaching method over another. The purpose of Descriptive Research is simple to show conditions as they exist. Correlation analyses, surveys, case studies, and direct observation are techniques used in descriptive research. (19, pp 35-37). Since it was neither practical nor possible to do more than a survey of the wide field of research during the Institutes, it was decided that the focus should be on descriptive research. This focus was selected because reported research in this category is often the most easily understood and the most applicable to the classroom. It was further recognized that because of the nature of the techniques used, the school offers a suitable environment for this type of inquiry.

The need for descriptive research in art education was most clearly emphasized by Robert L. Lathrop at the Seminar in Art Education for Research and Curriculum Development. He states:

> ... research in art has been too "spontaneous", attempting to cover the "canvas" with broad sweeping hypotheses and explanations before an adequate background of descriptive and rational data has been obtained. This is not to be critical of the researcher who plows headlong into a highly complex area such as teaching and learning...It would be unfortunate, however, for any field to have too many "plungers," for unless research in art education is fundamentally different from any other discipline, it is the plodding accumulation of carefully obtained descriptive and predictive data that underlies any major theoretical or conceptual breakthrough. (1, p. 348).

As the planning groups and the instructional staff began to work on the specifics of the program, it soon became apparent that even the area of descriptive research was too broad to cover in any depth. It was

the intent of the group from the beginning that the Institutes would not be simply a series of lectures about research methodologies but that the participants must have the opportunity to become actively involved. Specifically, it was intended that each participant would write a proposal that would utilize what he had learned to help solve a real problem he had encountered in his current teaching situation. With this in mind, the focus of the Institutes narrowed to the methodology of direct observation. Observation was selected as the place to start since it is the most fundamental of all the methods and is the foundation of all research.

It is the hope of those of us who have worked on these Institutes and this publication that the contents that follow will activate an interest in research and encourage art teachers at all levels to seek more information* and try this methodology in their own teaching environment.

* Appendix A

Observation: The Basis for Teaching and Research

Dale B. Harris

I'm making this presentation with two observations in mind. I've worked with many art educators over a period of years. I make these observations without value judgments, I trust; but they are implicit in what I try to do with your profession. First is the observation that many art educators distrust science generally; they believe, for one reason or another, that it attacks and destroys something valuable in your field. Second is the observation that many people in art tend to be, in personality, rather wholistic or gestaltish. They prefer to deal with wholes. This attitude consequently leads to a skepticism of or at least discomfort with science's most powerful methodologies--namely, analytical methodologies. What I am trying to do is based on these two observations which, as I say, I am not judging at all. They are general characteristics which we must take into account in any discussion of research in art education.

I am going to avoid technical terminology of research for the most part, though not the concepts, I trust. I cannot be entirely innocent of some technical terms, for they are necessary to establish meaning. But I will try to avoid jargon. Consequently, I will be selective, particularly with regard to observational method. My field is developmental psychology, particularly child development. Unlike much of experimental psychology, this field rests its work heavily on observational methods. I suppose this reason prompted your committee's bring me into this picture as possibly offering some help to you.

Can We Apply Science to Art?

Byway of introduction, let me attack directly the problem of art vs. science with a few comments.

There are some who see art and science as quite distinctly different approaches to reality. They feel that the systematic rationale and the analytic character of science is incompatible with the intuitive and experential character of art or the aesthetic. Therefore, it is really quite inappropriate to speak of scientific inquiry into art products or processes. Indeed, there are some who insist that there shouldn't ever be an attempt to look at art from the viewpoint of research.

There is another view, that art and science come very close together

at the point of the active discovery or the creative synthesis. The artist must integrate his product from a variety of skills and components. So must the scientist, who must reconstitute the phenomenon that he has broken down and observed, studies, and manipulated in detail. A very interesting book appeared some years ago by Geoffrey LaPage, an English medical artist, (14) in which the concluding essay put this point strongly and effectively--that art and science really do come together. Their goals are almost indistinguishable at the point of the creative act or the act of discovery; though their methodologies may be quite different at some points, they do have this essential aspect in common. It is true that science focuses on behavior and dismisses the ineffable. Some artists wish to argue that there is an ineffable component in art, something that you can't observe, that you can't scrutinize in any systematic way. This may be true, but it is beyond our present science. There does remain a great variety of behaviors that can be observed and treated by research technologies. Indeed, if we are to teach art, whether its techniques or its appreciation, we must make some systematic study of these areas. If you will grant this point (and I am perfectly willing that you not grant it, in which case you'll have to approach all that I say with a good deal of reservation) that there are some domains of art which are subject to systematic observation and treatment, and desirably so if you are engaged in teaching, then let's proceed.

Observation-the Basis of Science

All science starts with observation. The observation is that something occurred, or it did not. One characteristic, and some have said the essential characteristic of science, is its public character. What one scientist has observed by a particular method of study or procedure, another can also observe by adopting the same method or procedure. Individuals with similar techniques and similar concepts will arrive at identical or very close to identical results. This point is essential to science.

The public character of art is not so clearly apparent. In science it is an absolutely essential ingredient. It may not be so essential in art. We have said that art has an exceedingly idiosyncratic or highly individual component. We state this point in various ways, that art is a relative matter, that it is what the artist intends, or what he creates, or what art is in the eye of the beholder. Yet, there must be some public character to art, or the disciplines of art history and art criticism could not exist. It seems we must admit some "public character" in art also, even though we may argue, and to some degree successfully, that there is a strong individual component in art--that art is what the artist has created or art is what the beholder conceives it to be.

Basically, by observation we mean we can supply a description of something that has occurred. Someone has said that nothing is so durable as a good observation. That is, a well described protocol of something that has happened can be examined now, and it can be examined two years from now, because it preserves essentially the ingredients of what has happened. A good observation, then, possesses the public character we have said was essential to science. We recognize, of course, that it isn't always possible to have a full account of what happened. As one sits watching and listening to a lecturer, one can make notes which reconstitute for him the essence of what the lecturer has said. One could, also, if he wished, make some notes about the lecturer's habitual gestures, some of his failings in speech. Such notes would enlarge the description of the lecturer's behavior. One could make some inferences about the state of his health--that he seems to have a cold, but there would be a variety of internal processes that would not be available to the observer from his eyeball technique of watching--the lecturer's blood pressure and other physiological indices of his state of arousal or emotional disturbance at the moment, his involvement with what he has to say, his reasons for accepting the lecture assignment, and many other significant items.

Characteristics of Science and Art

One's observations are necessarily selective, partly in terms of what's available to his method of observation, and partly in terms of what one chooses to define as relevant to his purpose. The fact that this selection occurs is due to large part of one's training by sitting through many lectures. One trains himself to listen and take notes, to listen to and note the content, and to ignore or attend only in passing the jokes and trivialities. This selectivity in observation requires that we define what we wish to observe, in order to state whether it did or did not occur. In this case the subject is observational methods, the requirements of good observation, and some of the values of observations. The student's selectivity, however, tends to focus his attention on that something, and it screens out or turns off his attention to other somethings.

So it is with observation for scientific purpose. One is concerned with, let's say, the creative aspects of children's easel paintings in the kindergarten, finger paintings in the third grade, and the like. He must define what he means by the creative content or the creative expression involved and screen out or turn off some other aspects of those paintings for the time beging. One may decide that subject matter and color are all that he will attend to. This selectivity disturbs some teachers because they want somehow to capture all that's there. It offends me at times when I'm missing something, and I wish I could capture it because I'm focusing on a particular aspect. Yet I know of no solution--

one cannot deal with the whole and deal with it systematically except in the most general terms. There is simply too much there. One has to do a certain amount of narrowing and focussing. If something escapes this process, then regrettably it escapes. It may be a serious shortcoming of science that it can't deal with all aspects of the whole simultaneously. Any yet analytic research procedures have been very useful, and so we can't weigh either approach lightly.

The scientific approach to observation requires that it be systematic. One must use orderly, disciplined procedures such that he obtains characteristic examples of what he observes. That is, one can observe a child in art class, and talk about his interest in his work, describing a variety of behaviors that suggest he is highly interested. Or one can talk about a child that he thinks shows a good deal of promise in his work. Such discussion may lead to selection of products and the demonstration of activities that have led you to the conclusion that this is a child with much promise or talent. But what research requires is not just the selection of talented cases or the random anecdotes that indicate the child's interest, but the gathering of observations by a procedure that guarantees their representativeness. Now this is an essential concept, and I would not avoid a technical term. One must show that he has observed a fair sample of a larger variety of behaviors, and that he has selected a few, but not in such fashion that he bias his observations toward the best that the child can do, or the poorest, but a fair sample of his work. One does not want just the unusual or the eyecatching.

The Identification of Criteria

Having recognized that observation is selective but also representative, a very important point is that we must identify the criteria by which we do the selecting. Furthermore, we must identify the criteria by which we judge whether or not something has occurred. We may ask this question: Have these children produced works with imaginative quality? Or we may choose to speak of aesthetic quality. If we use the words imaginative quality, then we must define what we mean by image and imagination, and this gets us into psychological processes. If we study aesthetic quality, we must define aesthetic, and we will be involved with philosophic as well as psychologic issues. But we must define and set up criteria so we know whether or not the particular work that we've assembled from these children does or does not contain these qualities. If we are going to identify drawings or paintings that contain this element as opposed to those that do not, we must identify what they are. This identification requires a definition. In science, definitions are technical and stated mathematically. And if not stated mathematically, they are stated in words that have limited meaning. One of the reasons art educators get nervous when they use scientific procedures is

because the language of art is often highly idiosyncratic--even poetic in character. It doesn't lend itself to the precision that permits scientists A, B, and C to identify the same behaviors. This is a very real problem, because the language we have to use must shave away the highly idiosyncratic meanings and come up with the common meanings. In art we find ourselves frequently using similar words but with each his own private meanings.

For those aspects of artistic behavior that are subject to scientific behavior, then, we must work toward definitions. While I doubt that we'll get many definitions that are mathematical in character, we will have to come to some agreement about what we mean. To this problem, we do have a sort of answer, and I suggest that this answer can be important for us. I try to describe a quality I see in samples of art work, and finally I say: "I can't tell you, but I can show you an example of what I mean." I shuffle through a pile of paintings and pull one out. "Now this has what I am talking about," I say. This procedure of locating examples helps when one runs into problems of defining in words some component or quality of the work he is trying to identify and observe.

I don't know how many of you know the work that Florence Goodenough did in the 1920's and which I took up years later, using children's drawings to estimate intellectual maturity. (11 and 12) This scale depends on the number of ingredients which may or may not occur in the drawing. In revising her scale I tried to use visual examples of what would be acceptable and what would not be acceptable for many points on this scale. Sometimes the tests manual gives an example of a "marginal pass." Nothing less than the quality exhibited in the visual example will pass.

Another example comes to mind. A psychologist was interested in studying whether, in fact, psychotics in state hospitals produced paintings that were different from those made by normals. He listed a long series of ways in which such paintings might differ; they included such things as color and use of line and mass, type of brush strokes, and the amount of paint applied in a particular way. I suppose he was influenced by his reading about Van Gogh and what had happened in some of his work. He proceeded to set up examples for all the qualities he wished examined. When he prepared a verbal description of the quality "thick paint" as opposed to "thin paint," he gave an actual example of each. When he described the distinction between heavy brush strokes and light brush strokes, he gave examples, as well as the verbal description. In judging a painting, a person could look at examples of specific qualities and make a judgment as to whether this or that degree of the quality was present. He found that giving examples along with his verbal definitions improved greatly the agreement of judges with respect to the qualities judged.

Classification; the Application of Criteria

Having defined our criteria and having aided the definition by examples that exhibit it and those that do not, then our procedure is to take a set of materials and sort them into groups--those that have the qualities defined by the criteria and those that do not have the qualities. Having made this first judgment, which results in sorting materials into two piles, we may look through the piles much more carefully to find more precisely the ways in which these two groups differ. This technique requires us to refine the criteria and is an essential step in research on artistic behavior.

Usually we say research could go forward in two steps or stages. There should be trial work on students who are like the ones you are going to study in detail, but not the same subjects, because from your study of their drawing you are going back to change some of your definitions and criteria. Having got to the point that you are satisfied with your criteria, then you apply them to the study group proper. And you apply your criteria to <u>all</u> of the subjects that you've selected for study.

The reason why one must separate these two steps is that if one changes his criteria in process, he really must start over again to apply the changed criteria to a fresh group. Otherwise he contaminates his original study by the changes brought about in his criteria by the examination of actual materials. I can't emphasize this fact too much--that observational techniques cause us to refine and redefine our criteria. In my field of child development, colleagues who have tried to define their criteria in terms of theory alone, and then go out to observe children's behavior have never been too successful. The simple criteria defined in terms of one's own ideas and concepts do not reflect adequately what occurs in actual behavior. One needs to correct his own impressions by actual involvement with material. Try out your preliminary criteria and definitions on some material and then go back and correct your criteria in terms of what you've actually seen. Having arrived at a redefinition, it is equally important to go ahead and check out what one has redefined against a new group of subjects, to see systematically whether in fact the criteria now work on this new group.

The Identification of Variables

Earlier we pointed out that the first step is to determine if something had occurred or not. Having made this distinction, we sort work into two piles, those that contain the event and those that do not. Then we make more intensive studies to see what similarities and differences exist within the two piles of work, especially in the group wherein the event has occurred. We seek to extend our original observation and make it more precise, so to establish similarities and differences. In

distinguishing between drawings that have a creative quality and those which do not, we may note that the creative examples show many features which cause us to wish to sub-divide the general class. Thus we establish logical categories, and these categories refine our criteria.

The ascertainable variability in the quality we originally defined may take the simple form of amount rather than kind. We fine ourselves making a judgment of amount on the general phenomenon that we define. We may refine this quantitative judgment by creating three sub-categories--a just perceptible amount of the quality, considerable of the quality, maximal degree of the quality. We may not fully agree on specific instances as between two of the three sub-categories, but there should be some degree of agreement or we are not judging by the same criteria. When we have divided this category into three sub-categories of a quantitative character, we have, in effect said that we are dealing with a variable. Here is another technical term. Something that exists in varying degrees, and we have made a first step into quantifying that variable by this three-fold division. Saying little, some, much, is a very crude but nevertheless legitmate kind of measuring instrument or scale to place on a variable. As our measurement becomes increasingly precise, our categories may increase in number. We may discriminate five or seven amounts.

Theoretically, we could increase these categories indefinitely and have what we face so often in art education--as many categories as we have examples to begin with! But the reason we are attempting systematic observation is to reduce the array of material given us by children into a limited number of groups or categories that we can make general statements about them. This, of course, is difficult. At professional meetings how many papers concerned with children's drawings or art quickly come down to showing slides of individual children's drawings! These are, of course, interesting, but do not permit us to make any classification. We just look at a series of slides. The attempt to apply scientific approach to material of this kind requires that we reduce the infinity to a series of finite categories. Thus, the problem of categorization is really significant. What we're doing is reducing what was originally a qualitative feature into categories, so that brief descriptive statements can be made and summaries can be drawn about the phenomenon we're observing. We can't be satisfied by describing the quality as infinitely variable. We can't take 120 drawings and say "Here are 120 examples of behavior." We seek to reduce the variable to a series of manipulative categories so that we can make some generalizations about the 120 examples. We reduce our observations by making the process selective and by defining our criteria. Having done that, we establish the categories that implement our criteria and give them meaning. We then make generalizations in terms of those categories.

Oftentimes when we do this reductive process, we fail to retrace our steps and to put our results into our original terminology. One can illustrate this failure from some of the research articles in art education. These studies report statistics, averages, standard deviations, and correlations, for example. The statistical tables convey information (meaning) in much reduced form. These tables are quite meaningful to a person who understands statistics but not very meaningful to a person who may know a lot about art education but not about statistics. It seems to me important to go back from those tables to the kinds of verbal description which will help reestablish the richness of the original behaviors which these tables merely summarize. Statistical tables are shorthand notations of generalizations about the observations made under the categories that have been established.

The Analytic Method

Essentially, what we are talking about is the analytic method. We break a phenomenon down into bits, and classify, treating the parts systematically. When we look at parts of art products, the appearance is very different from the original material. One feels relieved, somehow, if he can show a few examples. The temptation is to return to an exhibition. Examples are excellent. Let the examples, however, be selective to illustrate the generalizations achieved by the reductive process. Remember that one is always dealing with variability and that one selects examples that illustrate certain classes defined on a continuum which varies by small amounts.

Two illustrations of research that have covered this ground may help. First, there is Florence Goodenough's scale for evaluating drawings of a man from which she derived a judgment of intelligence. (11) This scale illustrates the analytic method. Second, is Professor Rabinowitz's work on teacher behavior which illustrates the qualitative approach to drawings, making judgments which evaluate the drawing as a whole. (16)

An Example of a Quantitative Scale

Florence Goodenough started with certain observations about children's drawings. First, she knew that drawings grow more complex with age. She also knew that the drawings of brighter as compared with duller children show very obvious differences; particularly in the richness of detail and proportions of parts. She recognized what must be systematic in her observations. Whe decided she would set a common task for children in these terms: "Draw a man, the very best man you can. Draw me a whole man; not just the head and shoulders, but a whole man." This task immediately restricts the kind of observations she could make. She could not tell anything about the child's treatment of scenery, or houses; she couldn't tell much about his sense of beauty or much about his creativity. This restriction lost much that might have

been observed, but for her purpose, which was to see how drawing relates to the increase in intelligence, it was necessary for her to restrict her observations.

Goodenough's basic assumption was that intelligence grows with age. Her hypothesis was that because drawings increase in complexity and in the use of realistic proportion with age, there may be a relationship between these two variables such that drawing might be used as an index to intelligence. She made the further assumption that the brighter, young child is more like the average child who is older than he, and that the duller child is more like the average child younger than he--both in intelligence and in drawing. The assumptions in regard to intelligence had been well demonstrated by research on intelligence tests. The relationship of drawings to intelligence had not been demonstrated except in a very general way.

Dr. Goodenough collected several thousand drawings made by children of all grades in elementary schools in New Jersey. She first sorted them out by age and went into a big empty room, spreading the drawings on the floor. She then walked up and down the arrays of drawings noting down her impressions as to the ways in which the drawings changed with age. She got many, many impressions which became the working categories into which she then proceeded to classify the drawings. For example, she got the impression that, with age, children treated differently the shoulder of the man. Where younger children were not too precise about locating the arm on the body, the older children were quite particular. Older children were much more likely to include a variety of facial detail; young children were satisfied with the major features--eyes, nose and mouth. Older children were more inclined to put on items of clothing.

Her first criterion, then, became increase with age in the percent of children including each feature of detail. If there was a regular increase, with age, in the percentage of children including the feature, this item was retained for her scale. If the item did not show an increase, she discarded the item from further study. Then she went through the drawings and sorted them first by grade then by age within grade. Presumably, the children younger than the typical age for a grade were the brighter children who had been pushed ahead. In the 1920's acceleration was not uncommon. Persumably also children who were older than typical for a grade had been held back because of poor performance. Presumably these included many slow or dull children, intellectually speaking. This procedure supplied her second criterion, that the under-age, presumably brighter children should include the item in greater percentages than the over-age children in the same grade. If this second criterion was satisfied she kept the item. Some fifty-two items satisfied both criteria. These items were, you might

say, fifty-two categories in terms of which drawings could be divided--taken one at a time. She found that if she merely scored the drawings as a test, counting a plus for each item in her series of categories which appeared in a drawing and counting as zero those which did not appear, she got a score--a number of points that varied from child to child. She found that this score correlated statistically with other measures of intelligence which might take an hour to two hours to give. The drawing of man took, generally, about ten minutes. In the elementary school, she believed her hypothesis justified.

Drawings treated in this way gave a measure of intelligence, and this measure has proved to be quite a useful, simple measure of general intellectual maturity in elementary school children. My revision has amply substantiated her hypothesis. She used an analytic procedure to identify the component in drawing that relates to intellectual maturity. She was not concerned about other features of children's drawing or their drawing behavior or ability. She scored not on technical or artistic excellence, but on clear representation of the idea embodied in the item. For example, the item: child draws the eyebrow or the eyelashes, or both. The percent of children who include this point in each age group from six through fifteen in representative samples of one hundred at each age goes up steadily from 30% at age 6 to near 80% at age 13 and then drops slightly at ages 14 and 15. Adolescents, typically, become less certain about drawing ability and less willing to attempt precise details that small children do boldly, so the percentage curve drops a bit in these older ages. As a measure of intellectual maturity this item is no longer functioning so well at these ages.

An item which works adequately at the oldest ages is the point: feet shown in perspective--one or both feet in perspective. However at no age do very many of the children include this, no more than about 15%. In one sense it is not such a good item because it ought to be included by perhaps a third of the children to be functioning as a measuring device. We kept the item in the scale because we were trying very hard to find items that would differentiate among our older children. We were not very successful, and this was one of the few items that did, so we kept it.

Young children frequently draw teeth, and so we put that item into the scale. Children at six and at eight included it, but no one after age twelve included the item. Furthermore, while it differentiates the bright from the dull at age six in that the bright children are more likely than the dull to include it, at eight and ten it differentiates in reverse; the duller children are more likely to include it than the brighter children. The item, then, was discarded on two counts--age progression, and consistent differentiation between bright and dull.

These data illustrate an analytical procedure and how observation of drawings can be narrowed by definition to fifty-two elements scored on their presence or absence and treated systematically to yield a measure which has found considerable use in psychological clinics throughout the country since it was first introduced in the mid 1920's. It is a widely used test with young children because children like to do it. It doesn't take much time to give or score. It gives a reasonably reliable index for a rough measure. We don't use it when a precise measure of intelligence is required, because making a judgment about a child's school placement or his placement in an institution on the basis of intelligence is a serious matter, and we don't want to rest the decision on any one test, let alone a ten-minute test. We want the judgment of a variety of tests. This is just good psychological prudence. But the Goodenough method illustrates some of the things we've talked about--the analytical procedure, the discrimination of elements for observation, and the definition of those elements. Finally a point I haven't mentioned heretofore--the fact that you can train clerks with no more than high school education to score this test as competently as highly trained psychologists. That we can do this testifies to the public character of the phenomenon we have observed. People who follow the directions do not require a lot of special knowledge to agree. The point is explicit and objective. People can arrive at similar results after a short degree of training in using the scoring instructions.

An Example of a Qualitative Scale

The other research I wanted to mention was a study by Dr. Rabinowitz. His problem was to identify students in teacher training classes who had considerable promise as teachers. His criterion, ultimately, was their skill as classroom teachers. He tried to devise a measure, based on drawings which these student teachers made, which would predict certain qualities of successful teaching.(16) I'm not going to talk about the ability of this test to predict teaching performance but rather the construction of the scale, which illustrates the development of scales to measure qualitative aspects of drawings. He constructed a scale for treating the drawings supplied by students, so that he could make this other kind of predictive study. He asked students to draw a teacher teaching a class of children. He emphasized that this task need not require an artistic talent, that he could accept sketches however crude, so long as they illustrated the student's idea of a classroom. From a preliminary scrutiny of drawings he believed that he could identify five dimensions or criteria which could be evaluated in the drawings. One of these was the relative emphasis given the teacher in the drawing. The second was the initiative expressed by the teacher figure in the drawing. The third was the psychological distance between the teacher and the pupils, the fourth was traditionalism in classroom arrangement. The final dimension was artistic quality, which Dr. Rabinowitz put in as a control scale to make sure that when his judges evaluated the first

four elements they were not just evaluating some aspect of artistic quality in the drawings. It was necessary to demonstrate that when judges evaluated the drawings with respect to these categories, they were judging something other than just the pleasingness or the skill in the students' work from a representational competence point of view.

It was necessary to define these dimensions somewhat more extensively than by merely naming them. For example, Dr. Rabinowitz said one may judge relative teacher emphasis in terms of the detail of the teacher figure as compared to the pupil figure. If the teacher figure has more detail, then there is more teacher emphasis than pupil emphasis. Also, one could take into account the placement of teacher in the drawing; was the teacher figure "central" to the drawing? The heaviness of line on the teacher drawing as compared to the pupil drawing, and the relative size of the teacher figure as compared with pupil figures were two additional criteria by which teacher emphasis could be judged. These four criteria were as far as he got, verbally, in his definition of the dimension, "teacher emphasis." But he found he couldn't judge his drawings successfully just with these verbal guides alone, so he selected examples which expressed five degrees of "relative teacher emphasis," scaling the examples by a systematic judging technique. These examples represented about equal distance from "teacher minimally emphasized" to "teacher maximally emphasized," scaling the examples by a systematic judging technique. These examples represented about equal distances from "teacher minimally emphasized" to "teacher maximally emphasized." These examples have a qualitative character, but one can lay these five examples in front of him and take a pile of drawings on the same subject matter, making a judgment on each as to which of the five standard examples it most resembles, qualitatively, in terms of the defined dimension. The examples supplement the verbal statements of the criteria comprising the dimension. When you have computed an average for your sample, you may find that the average drawing by a particular group of teachers is 1.7; that doesn't say very much to you. So you go back to your scale and say that it lies between drawings numbered 1 and 2 on the scale, closer to 2 than to 1. This procedure brings your statistical shorthand back into the visual language you started with and makes it more "visual" perhaps.

Actually, when one makes this kind of evaluation of an observation, it is generally well to judge all drawings on one scale at a time because one wishes to keep clearly in mind the dimension and its criteria. You don't wish the judgments on one scale (or dimension) to color the judgments on another. Since he used the same judges, Professor Rabinowitz allowed several days to go by before judging the next dimension in order to erase the memory effects of the previous judgments. Thus he sought to reduce the "contamination effect" in judgments.

The psychological distance scale was defined by two verbal criteria, barriers (such as desks, etc.) interposed between the teacher and the children, the accessibility of the teacher to the children as inferred from the gestures represented in the teacher or the children figures in the drawing. From an arm-chair point of view one might suppose this dimension somewhat harder to judge than the teacher-emphasis dimension. But in fact his judges reached a more consistent degree of agreement on the psychological distance scale than on any of the others. The "traditionalism in classroom arrangement" judgment might seem to be somewhat easier to judge; yet the judges agreed only to the extent expressed by a correlation coefficient of .53 and was the poorest scale from the standpoint of the public character of the judgment required. The elements or criteria of this scale included pupil grouping (is the class in a single or multiple units), the nature of the seating, the reliance on furniture to designate a classroom as such, and the existence of special interest areas in the classroom portrayed.

Many art educators find the quality scale more acceptable than the point scale because it preserves the totality of the drawing more effectively. Many times as I have worked with Goodenough's method, the analytic method, I have felt frustrated by the fact that there is a lot in drawing that I lose by attending only to the defined points. Then I have to remind myself that it is quite all right for the purpose for which the scale was designed. The details were selected to define the dimension of intellectual maturity. Of course there are many features of the drawing which may tell other things about those who draw them, and the point scale technique, as defined, misses them. A quality scale can often create the impression that one is retaining more and losing less, but again, this depends on what criterion has been selected, and how carefully it has been defined.

In the Rabinowitz study the artistic quality scale correlated very slightly with each of the other four; in fact his judges were judging something other than artistic quality, and it is important to know this. This procedure established, by a kind of elimination process, the fact that he was measuring <u>something</u>. The other four scales intercorrelated quite modestly, arguing that they did not merely duplicate each other. They were measuring reasonably independent dimensions of whatever it was he was measuring. Unfortunately, the research wasn't completed, by correlating scores on these scales with measures of teaching proficiency achieved in student teaching or, later on, in professional careers. Although we do not know what the predictive power of these measures was (the purpose for which they were originally intended), we do have a clear demonstration that it's possible to construct scales which measure certain conceptual dimension of behavior and to do so consistently. This is the first step. The second step is to determine what the scales do measure. That, of course, is crucial also. But if

you cannot measure something reliably and consistently it is perhaps foolish to proceed further. One must demonstrate that his definitions have a public character.

Summary

Thus, the first step in scientific work is to determine what you wish to observe. This requires setting up some criterion. Usually, this procedure first requires a definition and then the identification of examples. Having determined which examples illustrate the presence of the quality you're observing, and which represent its absence, the second step is to determine additional sub-divisions or categories, varying in amount or kind. In discriminating such subclasses one further defines his criterion, sharpening it, making it more workable. Then having demonstrated a workable criterion, carefully defined verbally and pictorially, one can give it to someone else. If a second person can make similar discriminations, one has not perceived something worth observing, or he has been unsuccessful in defining it so that he can communicate its character to others. One may think that artistic merit can be judged, but if someone else similar in background and orientation cannot use a definition of artistic merit to reach reasonably similar results, then something is wrong from the point of view of science because science insists on the public charactor of information, on the replicability of results.

Varieties of Observation

Dale B. Harris

In the previous section we started with the notion of something we could observe. We asked the basic question--has something happened or has it not? Observed in this way we sometimes speak of a characteristic as an attribute--it exists or it doesn't. We may divide people into two groups according to an attribute--those who possess a quality and those who do not. A person has curly hair or straight hair. We can, sometime, break down attributes into classes based on more than just presence or absence of one quality. We may decide that "curly" hair breaks down into three sub-classes--wavy, curly, and kinky. Another example, one might take a pile of children's drawings (when they have been free to select their own subject) and create a classification as an extension of the attribute system, according to the subject matter portrayed: landscapes, human figures, machines (automobiles and vehicles of various kinds), etc. Such a classification is sometimes called a nominal scale; one classifies objects into a discontinuous series or categories by naming. For a nominal scale there is no intrinsic or necessary arrangement or order to the categories that comprise it.

However, we may go a bit further. The minute we say that something exists, we can very often say that it exists in some amount. The moment we try to define "amount," we're talking of a variable. An attribute which can take any one of a number of quantitative values is a variable. In such case, we set up a set of categories by amount, those that have none, one, two, three, four, or more of this characteristic or , in more general terms, none, some, considerably, very much, and extreme degrees of the quality. As in the example of an attribute which can be classified nominally, so we may classify into categories of amount, either counted or estimated. But now we can see a sensible order to our categories, from least to most. We now speak of an ordinal scale. Nominal classes can be arranged in any way we want without destroying the characteristic that we're classifying. Ordinal classes can be arranged in only one way, or we destroy something of the characteristic that we're classifying. An ordinal scale for variables can take, and often does, take two forms--qualitative and quantitative. In the quality scale the categories are arranged in order, but the steps are defined by examples. The intervening steps or degrees are not estimated; therefore the scale,

though ordinal, remains qualitative. The amount aspect of the scale is reached by estimating. A quantitative scale is illustrated by the Goodenough Drawing Test. (11) The number of points scored on it may be counted to yield a quantitative measure. The scale is ordered from 0 to 52 points. One may group these units into classes of several units (3, or 4 or 5) and reduce the number of groups into which one would sort his results. But technically the scale runs from 0 to 52. One must distinguish, then, between the quality scale in which one makes a judgment and the quantity scale in which he counts or measures with a quantitative scale.

The Idea of Dimension

Fundamentally, when we construct scales we are dealing with dimensions. A dimension is some quality or aspect or property of an experience or of an object which can be defined and which takes different values or degrees. When we discriminate among these values or degrees, we define a continuum. This continuum we see as a variable. Furthermore, this continuum or variable must be pure, or unidimensional, in the sense that it deals with only one clear-cut feature of the quality that we are dealing with. When one takes a collection of children's paintings and asks himself how he may describe these, he must first determine possible dimensions in terms of which he might describe the paintings. Examples might be, subject matter, color, clarity or "muddiness" of colors, quality of line, use of space, use of mass, etc. One art group identified more than fifty features which might be considered as dimensions of drawings. Now, that is a lot of dimensions to work with, yet everyone of them could, theoretically, yield a continuum along which qualitative or quantitative categories or classes could be defined. Taken altogether these dimensions would describe all features of the set of drawings that this group examined.

Thus a dimension is equivalent to a continuum, to a variable, and this dimension is defined by the one common property that runs throughout all examples or degrees. If you mix two or three properties into your definition of a dimension, then you're in trouble. For one thing, the reliability of your judgment goes down. You seek to keep one property well defined, with the gradations or the categories established unidimensionally, and then you develop as many different dimensions as you need to adequately describe the material you're working with.

Dimensionality has one other feature which should come into our thinking. Dimensions may exist at various levels of abstraction, from very specific, and concrete, levels to very abstract levels. You may have a dimension such as amount of pressure exerted on the pencil; this is very concrete. You may have a dimension which speaks of the

character of the space enclosed by the line, which you then describe in some way. Such a dimension is at a more abstract level. It's dealing with a broader kind of phenomenon than just pencil pressure. When you get up such concepts as aesthetic quality, you're dealing with even a broader feature of the material.

When dimensions vary from the very concrete to the very abstract, we can often see a kind of order among them. We may then speak of a hierarchy of dimensions or of meanings. For example, you can have a set of categories which you call fruit. This concept creates a set of disparate categories which constitute a nominal scale because these are apples, and pears, and plums, etc.--all distinctive, separate kinds of fruit. Similarly, you may identify many categories of vegetables. Fruits and vegetables have certain features in common--both are perishable plant products, seasonal in character. Both are edible and found near each other in markets. Each is a class term and in a sense they are coordinate. These and many other items arrange themselves into broader and broader categories, until we speak of "food"--a general term which embraces a hierarchy of terms. A hierarchy is built from concrete to more remote levels of abstract concepts. Dimensions have this property of arranging themselves along degrees of abstraction into a hierarchy. This feature of dimension is well to keep in mind when you're building quality scales. You may be able to arrange the dimensions you identify into a hierarchy. If so, you have taken a significant step toward establishing meaningful relations among them.

To summarize, one fundamental aspect of research is the defining of dimensions, and establishing categories within them to permit one to classify observations in appropriate ways, so that general statements can be made. This is, one seeks to reduce the total number of classes from N = the number of subjects or examples of work you're working with, to a much more limited number of groups. Each group or class will exhibit some intrinsic characteristic that holds it together, that makes it a class, that defines it, and at the same time, permits you to observe the regularities or continuities among these classes. From this process of classification of observed phenomena you can make certain generalizations about the material. That is, you can divde your children's drawings by age of artist and by subject matter. Then you can make general statements such as "boys are more likely to do ships than girls"; "primary children are more likely to draw flowers and birds than intermediate children," etc. Or you may work at a more abstract level and define dimensions which show use of line--whether and how lines enclose space, etc. You then say children at the primary level are unlikely to make abstract forms spontaneously, or they're very unlikely to use perspective. One makes such general statements after he has defined the property he is looking for, sorted his materials into the categories, and made his estimates or his counts.

Identifying Meaningful Dimensions

The kind of dimensions and the categories that you define within these dimensions will depend on the scope of the task you've initially given your subjects. Again, I'm using as illustration the drawing area. If your instructions allow children to draw anything they like, you'll need a much broader set of categories than if you say "draw a man," because you can get so many different kinds of drawings. To make dimensions manageable in your research, you may have to set narrower limits thant "draw anything you choose." Your task is considerably more manageable if you say "draw a picture of your school yard at recess time." Here you've limited the subject matter, so the number of categories you must establish for classification becomes specific and thus more manageable.

The question should be raised: "How do you know what categories are going to be important?" "What dimensions, among all possible ones, should be studied?" Material may be treated in an infinite number of ways. What you will observe, and dimensionalize, will depend on your problem, your question, and on your previous experience and your knowledge of previous research. Sometimes the selection of fruitful dimensions depends on hunch. Scientists in many fields use this more often than they admit. We sometimes call this factor "intuition," and psychologists are sometimes inclined to toss it aside as mystical. Yet, I'm convinced of two things: first, intuition is not completely divorced from training. A person who knows a field very well, is going to function intuitively more adequately than one who doesn't know that field very well. If you're going to do research with children's drawings, you'd better have a lot of experience with children's drawings. If you do, you will select dimensions more knowledgeably than if you come at them cold. If you've handled lots of children's drawings, looked at lots of them, you'll have a better informed and successful intuition in regard to the kinds of dimensions you set up to classify them on.

The other thing I would say is that people do differ in this intuitive capacity, and beyond that I can't say more. Even the informed and experienced people differ; some pick fruitful hunches more often than others. This feature of intellect remains something of a mystery. We don't understand it nearly as well as we understand problem solving.

The General Observation

There are a number of ways observations can be made. One of these is certainly direct inspection and description. I'll just call this "eyeballing." One of the time-honored techniques in child study has been the running description of behavior from watching what happens. Much can be learned from general observation, usually that behavior is much

more comples and difficult to describe than one assumed. This is an important lesson to learn. Occasionally I get experienced teachers taking my child psychology course. I send them out to observe and record everything a child does in twenty minutes of free play. They often object to this, and say, "but I understand children; I've taught for ten years." After a certain amount of protest, they'll react that "I did'nt know children could do so much, particularly young children." "I found myself hard-pressed to put down everything that went on in twenty minutes even in general terms." That's one discovery from general observation.

Next I ask that the observers do one other thing--separate what happened, the behavior, from their own inferences and interpretations, which are to be noted in the margin of the behavioral observation record. The student writes "The nursery school teacher took the child by the hand and led him, unwillingly, over to the sandpile." And I say, "Well now, wait a minute--the teacher took the child by the hand and led him to the sandpile. Unwillingly was your inference. What led you to say unwillingly?" "Well, the child cried a little bit, or held back." "Well, that's behavior; write out that observation as behavior. The unwilling is your interpretation and should be noted in the margin as an interpretation of behavior observed." And we go through this process over and over--it takes a bit of doing--to separate the inferences from the behavior.

By the time you get people separating inference from behavior reasonably well, you'll find that they make a couple of discoveries, that it's very difficult to make strict behavioral accounts and get everything down, because behavior is very rich. Very often students will say, "Why don't you just send me out with a movie camera? It would be more successful than my trying to write all these notes." The second discovery is that when all the interpretations are sorted out, behavior descriptions become very dull. Indeed the interpretations, the descriptives--"big," and "little," etc., adverbs "willingly," "happily," etc. make the behavior descriptions interesting. When you pull them out and put them among the interpretations, the behavior descriptions become very dull.

Observing the Stream of Behavior

A very extensive set of researches on the observational method, done by Roger Barker at the University of Kansas, (5 and 6) has deliberately reintroduced the adverbs into running accounts of child behavior because he finds that straight behavioristic records don't give him very useful research material, psychologically speaking. You can't observe a child's intentions or goals; you infer them. And the minute you pull goal directed behavior out of your observations, your observations become fairly pointless.

Barker built his whole set of observations around sequences of behavior, which he separated into episodes whenever the child's goal or action changed. Bill and John were playing in the sandlot with some old crates. The behavior record goes on to make many observations relating to their handling of material, skill, and problems they met and solved, etc. The boys were building a shack, but they dropped this behavior, got on their bikes and rode down the street. At this point their inferred goal changes. They went down to the drugstore for some ice cream. First the goal was to construct a hut; now the goal is to get ice creams. This change of behavior represents a way to unitize the behavior. Barker's whole object was to treat the stream of behavior as a stream, and yet, break it up into units that could be handled psychologically--categories, if you like. He found that he could do this by inferring the child's purpose; indeed, he had to infer purpose to study behavior in context meaningfully. He could evaluate these units of behavior in a great variety of ways. For example, he could evaluate a child's behavior in terms of the apparent clarity of the goal set for the child when tasks were assigned him. He could evaluate the amount of frustration experienced in sequences of episodes. He could sort episodes into social and non-social, and study them in new ways. Many interesting facts came to light when he studied behavior episodes across many children and across many episodes. He found for example, that the clearest goals for children are set, not by teachers, not by parents, but other children. Teachers actually are least clear in making assignments and setting goals for children in the elementary schools. Parents are a little better in setting clear goals than teachers, but not much.

One may get the impression from some of the clinical literature about children that their lives are full of disappointments and frustrations, being put down by adults, etc. Barker found that if you really take the ongoing stream of behavior as it occurs in school, on the playground, at home, etc., break it into episodes, you may evaluate them against a variety of dimensions by fitting the behavior against rating scales, one scale to each dimension. By this method, Barker found that a child's rough experiences are very minimal. Acute frustrations occur infrequently. A lot of a child's experience is kind of pointless, but there is much more constructive than destructive in the child's daily routine. Adults, teachers, and children offer more help than hinderance. Barker was able to also take the school episodes, along with the other episodes, and classify them according to the learning significance of these episodes as they occur--did an episode seem to have a significant impact on the child, change his behavior? We find that more learning occurs out of school by far than occurs in school, and that significant learning episodes in school per child may occur as seldom as once in forty minutes. Mostly, the child is just sitting there. Well, we've known this in a general way, and we haven't been very happy to face it. Data like these bring one up short.

Making Judgments in General Observation

The straight behavioristic, observational account is, then, not necessarily the most useful. One may learn a lot about observation by forcing himself to record only behavioral descriptions. However, subjective or experiential aspects may also be observed. Such observations often upset some of our cherished preconceptions. Such data give us quite a different picture of the child's life from what our preconceptions would have us believe. These running accounts that I'm talking about can be done in free situations or in controlled settings. You can do them, for example, by giving children certain materials to work with and observing what they do with these materials. Or you may give them access to a great variety of material and see what they select. You may give a specific assignment, saying "accomplish this." Immediately, depending whether you've set up a very free or controlled situation, you require different kinds of categories. Once you have gotten the materials together, you classify them on appropriate dimensions or variables, assigning them to categories within each variable or dimension. In other words, you scale them. You may have 54 dimensions that you've identified as possible ways of evaluating a material, but you may find that in fact you have only 8 or 10. This is a common experience. I've heard Professor Beittel comment on this fact frequently, that one can think of many more ways to handle a material than you actually, psychologically, can use. Scale will be intercorrelated, some of them intercorrelated so highly that there is no sense in making separate judgments. Although the scales may utilize different words they in fact require essentially the same kind of judgment. This problem runs all through the descriptive terms that you use in evaluating art products. Very often our meanings overlap--we think they're distinct, but when we come to make the several judgments, we find we're judging fundamentally the same thing. And so in research we have to reduce the number of scales we apply to the distinctly different judgments that we're operationally making.

The Selective Observation

We have discussed the running account of behavior. There is also the observation of behavior in a selected situation. These situations are selected according to a dimension--an activity or product dimension. The observation is limited in time and the kind of situation you have selected in which to observe behavior. The resulting record can be subjected to rating. This observation also can be a training device--to make one sensitive to the distinctions between judgment and inference and what actually occurred.

For research one is much more likely to use selected situations than the running account of behavior. However, the general running account

method often can be helpful in defining your problem and determining what specific observations will be useful. For example, if your problem is the general one of how may we free-up children's drawings, the best thing to do is to watch children draw for awhile. Then as you define your problem more precisely you can determine the observations that you can treat more systematically. You select some drawings which will help you establish categories expressing degrees of freedom or constraint. Then you begin to think of what you can do as a teacher to increase or to decrease this dimension.

At some point you may wish to include in your observations a recorded statement by the subject. Now this gets us into interviews and questionnaires. The subject reports on his own behavior as directed or structured by the questions put to him--eigther written or oral questions. This type of observation is useful for aspects of behavior which are not easily observable; for example, a child's preference for one kind of medium as contrasted with another. You could get some idea of preference observing the child's behavior in spontaneous situations over a long period of time, or in free choice situations where, for example, you might note that he usually selects crayons over finger paints. But why not ask the child directly what he'd prefer? This is a legitimate approach. You can get choice information economically simply by asking.

A questionnaire or interview requires careful preparation in phrasing questions which will elicit the observation and information you want. This is harder to do than you may think. No one should ever sit down and write out a questionnaire or interview and then go out and gather data. One always must pretest his questions--sometimes several times. A good example is the Eisner's Art Interest Scale.(7) This is a schedule of questions which can be used to collect information about interest in art or previous experience with art.

This direct approach, too, is important in appreciation studies and in process studies. What you're getting at is the inner psychological processes that are not available to direct observation. You can say, "Well, technically, maybe they're not even available to the subject himself; maybe these are unconscious phenomena, and he can't tell you in words what's going on." This is a possibility. There are undoubtedly some aspects of appreciation that are not available to a subject consciously. Maybe you'll have to be particularly clever in devising some indirect methods of observation. Some psychologists have been particularly ingenious at getting indirect measures. But increasingly, as your measure becomes more indirect, you have more and more difficulty defending its validity. The questionnaire, or check list, or similar device is most legitimate for those areas of experience in which you have reason to believe a person can make some judgment on his own behavior.

Tests and Work Samples

Then there are tests or work samples. I'm talking about art products, the products of art behaviors; I'm not talking about personality tests. I am talking about tests of skills in art, or work samples of skills in art. Give a person some material and see how he works with it. How does he handle his brush with oils, with watercolor? How does he handle clay? If you set up a test or work sample, you assume that there's a body of skills that you can sample systematically. You ask a person to perform the tasks that you've sampled, and record his observed performance on each skill or operation. You accumulate these into sums of scores or possibly a profile. What you're really getting at here is observation of technical proficiency. Probably this approach would be used in art school much more than in public school art education where, as I understand it, your emphasis is less on technical proficiency than on other purposes. But, depending on your educational objectives, if you were trying to embrace or develop technical skill with a particular medium in the sixth grade, it would be perfectly legitmate to build a set of tests that would permit you to make observations about the child's technical skill. Finally, we should note that in all research the descriptive task is required at some piont. In constructing standardized tests, a great deal of observation goes into the work. Someone may say, "I'm going to use Elliot Eisner's Art Interest Test with a group of high school freshmen, and divide them into those with high interest and low interest and then study their behavior in drawing and sculpting." Well, the direct observation you might make with respect to art interests has already been done for you by Professor Eisner. In constructing his test he put into it a great deal of direct observational work which you accept as an accomplished fact. You take the test as a developed instrument and begin your observation at another point. Having used the test and sorted subjects into two groups, then you observe their behavior in art class. You take the test for granted. I merely make the point that you can't undertake any research without embodying observation. When you give tests, you're accepting somebody else's observations.

The Case Study

Dale B. Harris and Kenneth R. Beittel

A. The Case Method in Art Education

Dale B. Harris

There is a particular use of the observational method which has considerable application in art. The case study should have considerable appeal to art educators because there is a minimum of quantitative analysis involved. It is globalistic, and permits the synthesis of materials in general statements about one person, or one production. Those artists who dislike breaking a phenomenon into elements or into numbers may feel they stay closer to the significant material by using the case method. Many art educators deal with groups of children, and they may believe that the case method is not practical for them. However, it may be useful for particular studies, even when one is concerned with groups. The group process itself can be made the subject of case study.

Broadly considered, a case study consists of the collection of a wide variety of data and information about a single person, group, event, or process. One seeks to put together a complete picture, from which he may gain greater understanding of that person. Frequently, cases are studied collectively. That is, one gathers a series of cases, collecting similar data on all of them. He then reviews the cases to arrive at some general statement formulated from the consistencies or trends he finds among all of them.

The Case Method Gives Particular Information

If one wishes to describe the status of a group of children from a collection of individual case studies, he is not likely to get good information by this procedure. That is, one cannot get information concerning parameters or norms. One can't really tell what any population of children is like, because one rarely selects cases to represent a population. For example, in the example Professor Beittel presents, he advertised on the student bulletin board for subjects who would be willing to draw for an hour a week over a period of ten weeks. He had no control over the selection of subjects except

among those who came in response to his advertisement. He got some very interesting people, including a math major, a science major, an English major, and so on, but it's clear that this group did not constitute a random or even representative selection of Penn State students. They were students who were curious about or interested in drawing, or they wouldn't turn up as volunteers. Such cases are self selected; they are not a random set. Such students give useful research information for Professor Beittel's purpose--to study process in drawing. But they do not give a picture of what drawing behavior by the typical Penn State student is like.

A Study of Drawing Process

Nevertheless, by making such intensive case studies, of drawing process, the investigator can get a very clear picture of how individuals go about working--the strategies and the imagery they utilize--by getting thoroughly acquainted with these individuals in depth. For this purpose, except for skilled artists, almost any student will do. Several case studies will give Professor Beittel some idea about variability and constancy in process. Thus, for process information, for knowledge of a person in depth, for forming hypotheses about how something is accomplished, the case study is a powerful exploratory device.

Other Uses of the Case Method

One may make applications of the method other than that described by Professor Beittel. He can collect samples of art work of a particular individual over a period of time. One may keep notes on the comments made by one child artist while he works, notes concerning his apparent motivation or his method of work or the like. One may keep notes on classroom behavior that he feels throws light on one particular child's interest, personality, and the like. He may ask the child to write an essay or to make a statement about his interest in and his previous experiences with art; he may ask the child to write about a particular project he has done. One can economically collect information regarding prior art experience, by an inquiry sheet, to be included with the case notes.

A teacher can keep notes concerning a significant art experience he introduces in class; he may, for example, introduce some intense sensory experience, moving music, or tactual exploration of a variety of textures and surfaces, and then ask the children to draw or paint or model. He may keep notes of what children say and do, as well as their resulting products. Thus one may make "case studies" of groups or of individuals.

What are the purposes for which one might accumulate such data? One might be interested in accumulating records of individuals that he feels are particularly creative--imaginative people in an artistic sense. You just want to accumulate these for a time to see what you get. In this way I've been keeping notes rather casually on some exceptionally gifted children I run into from time to time. I've not done it particularly for research; I do it for teaching purposes, to have some specific illustrations for lectures.

You may wish to study individuals who have a particularly strong imagery or a particular kind of imagery. If you follow the late Professor Lowenfeld's tradition, you might want to study haptically inclined individuals, because such persons do not come along frequently in our visually oriented culture. You would keep notes on such individuals until you've built up a series of cases giving you a better understanding of this dimension of perception.

Studies of Growth in Interest or Skill

One of the best uses of the case study is simply to collect longitudinal records or samples. We have relatively few collections of such. Helga Eng has published two books on her niece's work.(8 and 9) Harold McCurdy, a psychologist in North Carolina, made a study of the drawings of a man (15) by a boy over a period of about five years. There's a published monograph concerning the work of a boy who was particularly interested in drawing trains.(13) A former student of mine in Japan has a remarkable collection of the art work of his four young children. He has kept everything that they have constructed, painted or drawn, including full notes of what they said at the time. He has an enormous pile of material which I have begged him to photograph in order to preserve, as this kind of collection is most unusual. Children's products are seldom saved systematically.

An Example of a Longitudinal Study

All these are case studies and may help give an understanding of artistic processes when we have a store of information in depth. I have some data collected in South America from children who had never used paper and pencil. Indeed they had never seen pictures. These children lived in a remote, isolated area, high in the Andes. The two cases presented here are by a four-year-old boy and his five-year-old-brother. I had these boys draw daily for about seven weeks. The samples included here were taken at intervals from the entire series. The work of the four-year-old is particularly interesting because the collection includes the transition from scribble to representative drawing. We showed him what a pencil would do and asked him to fill the page with pencil work, and this is what he did (Fig. 1).

The spiral scribble was our demonstration of what a pencil would do. Thereafter, our instruction was always to draw an object, generally a man. Fig. 2 shows his work two days later. Five days after the initial drawing we do get a rudimentary man, along with scribbles (Fig. 3). Two days after that the scribbling is much reduced (Fig. 4), but shortly reappears (Fig. 5), and is still present two weeks after the initial drawing (Fig. 6). After another week we have several forms (Fig. 7) with minimal scribbling. Note that the transition from scribble to figure drawing did not come all at once, nor is there much evidence of a progressive exploration of various forms of scribbling.

The following September, I visited the village again. Here is the first drawing (Fig. 8) that this little fellow produced, after six months of no drawing. The following day we get a much better form (Fig. 9). His five-year-old brother also shows a remarkable progression but without the definite transition from scribble. Fig 10 reproduces his very first pencil work. Again, the spiral figure at the top constituted our demonstration of the pencil. The following day there is a crude "man," along with some scribble efforts (Fig. 11). A week later all scribbles have disappeared in favor of forms (Fig. 12). Figs. 13 through 15 were selected at weekly intervals from the total series, produced daily. Fig. 16 presents this boy's first man after six months of no practice, and Fig. 17 was produced a week later.

From such case studies I can't get statistics, but one does get some appreciation and understanding of the drawing process and its development, and a vivid illustration of retention and, indeed, development during an interval of no practice. I must stress that these children had no paper or pencil in the interval. For these case studies we had no additional information about the children; it was just not available. They came from a visually impoverished experience; there were virtually no man-made forms except huts. There were no trees--only rocks and scrubby grass and llamas and sheep. Although there was no previous experience with printed pictures or figure drawing, yet from the start in the older boy, we got recognizable form, and in the case of the younger boy, a study in the transition from scribble to representation.

I recommend to the art educator the case method. Even though you work with groups, you can collect case material on particular children. Even though you're very busy, you can usually find time to collect data for one case in which you happen to be particularly interested. An accumulation of such over a period of years can lead to process understanding that group studies may conceal, as the following paper by Professor Beittel clearly reveals.

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

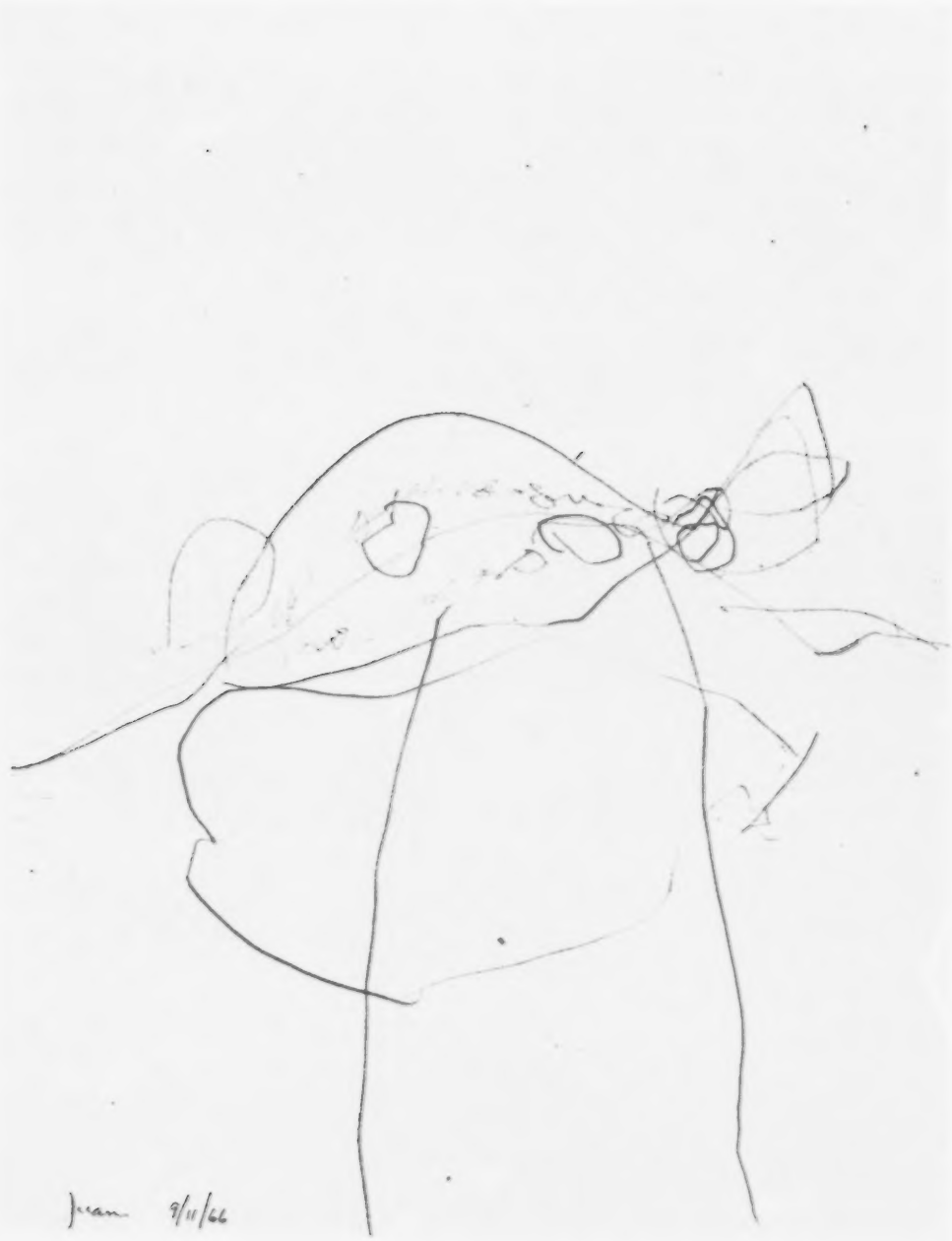

Fig. 8

Fig. 9

Fig. 10

Fig. 11

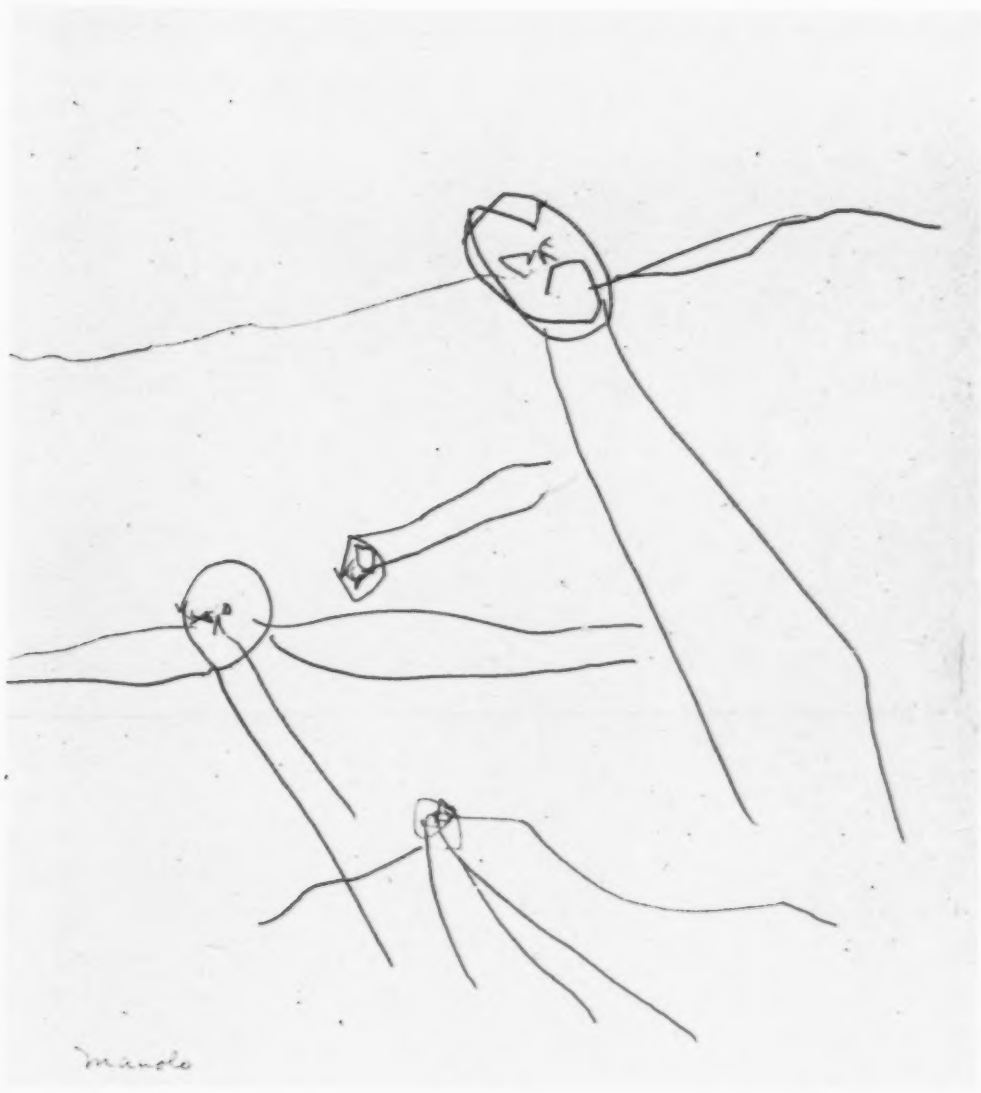

Fig. 12

Fig. 13

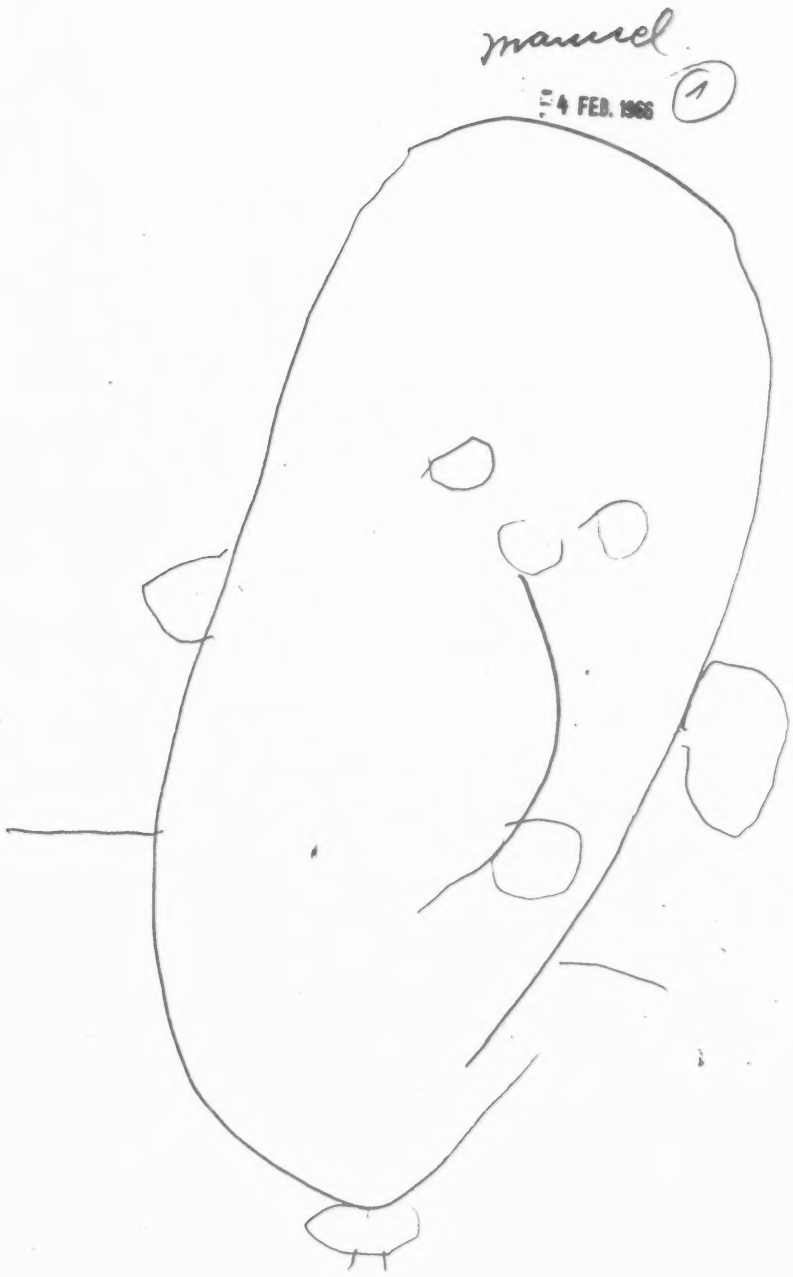

Fig. 14

Manolo

10 FEB. 1986

Fig. 15

Fig. 16

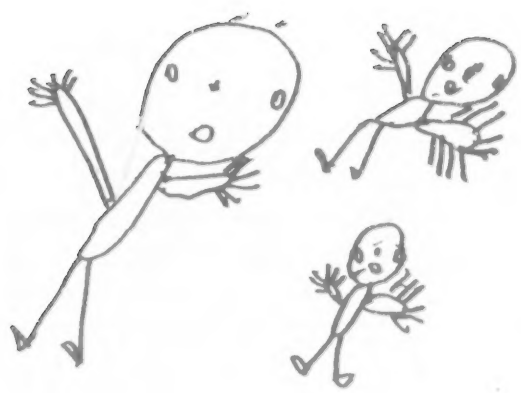

Fig. 17

B. Some Values of Case Studies*

Kenneth R. Beittel

There are several points of view one may take toward case studies. Sensitive experimentalists will frequently study single cases in some depth in order to develop an intuitive feel for phenomena they will eventually have to represent by means of concrete operations and well-defined, quantitative variables. Others will immerse themselves in the center of the many variables encountered in "real life" or "in the field" in order to develop their hunches concerning which variables merit further, more controlled studies. These are uses which serve a preparatory function. The case study is a stepping-stone toward research or a means whereby sensitivity toward phenomena is increased.

I have no quarrel with such uses, but I do think the attitude toward the case study under such circumstances is one which relegates it to an inferior position for knowledge-getting. I will assume that there are other uses of case studies which center value more within them, not on their auxiliary functions.

The forming of art can be seen as a process requiring an integration of socialized and idiosyncratic or personal, less ego-centered forces. Regularities in style, tradition, symbolism, and medium usage belong to the socialized side and submit gracefully to modes of inquiry common to the behavioral sciences. I have been able, for example, to study characteristics of drawing strategies (with factors influencing them) of defined sample groups. The noticeable qualitative differences between drawings of an individual, however, soon lead one away from the group focus and the variables and modes of analysis appropriate thereto. The socialized aspects do not cease. Rather the idiosyncratic ones press to the fore. More than that, it is the integration of these in the unique person which is the chief seduction of the case method for me. It is my feeling that the aesthetic and creative aspects of art-forming do not submit to generic viewpoints or criteria, but that they are experiential realities dependent upon the flow of articulation processes and upon the contexts in which these are embedded. This is

---

* The reader is urged to consult Gordon W. Allport's, The Use of Personal Documents in Psychological Science (3) for an extended discussion of case methodology which is admirable in its scope and balance.

a philosophical assumption, I realize, so I will not argue it here but merely state it.

There is a practical reason as well as a philosophical one for the study of the single case in art. The teacher assumes that his experience is a sound basis for interpreting the student's artistic processes and products. He arrives at a necessary humility in discovering how little he can legitimately interpret of another's art processes. The effort of nonjudgmental interest and attention required to focus on the wholeness of a case engenders the empathy and insight which cancels out his categorical expectations and preconceptions. He perceives that student as a lawful entity whose images, intentions, skills and circumstances are highly relevant to what he does in art. This might be called an aesthetic or appreciative function of the case method: the grasp of wholeness, origin, and continuity, of existential being and transcendence circulating around the continuous making of art. Further, the teacher learns to what degree his assigned problems and curriculum may be impositions on a student, deflecting rather than helping him. Still further, the teacher may arrive at a role much like that of the therapist (although not oriented toward therapy) where he feels himself mysteriously like a participant in a human dialogue which helps the other person direct his art dialogue in ways more satisfying to him.

To operate thus in the study of the single case and yet be a teacher may cause some confusion of roles. The observer whose chief interest is that of understanding the case is likely to be in the better position. On the other hand, the teacher who engages in such study may well modify his teaching methods to the place where there is little incongruence between the two roles. I have seen art teachers work effectively in this way, and there is a long-standing tradition in art education based on a belief in the organic "unfolding" of a person's expressive and artistic powers. It is not chance that Lowenfeld and Schaeffer-Simmern, for example, have often included individual case histories in their writings. By so doing they thought to present subtleties beyond their generalizations through a more or less "ideal type" which their generalizations could then refer to. But I have found, in my in-depth work with individuals producing an extended series of drawings, no uninteresting types and no ideal types. Rather the full flavor of uniqueness is experienced, along with the realization that all generalizations concerning a case are but tentative explanations and projections of the observer.

The case method in art, thus, leads one toward a phenomenological stance, toward philosophical questions about the experience of art, and away from group analyses and the kinds of manipulations more common to traditional behavioral science. It reveals to the teacher the way his students experience art, and it sensitizes the researcher to the full

range and complexity of the phenomena he studies. Case histories, moreover, can be structured so that they yield protocals which can be used in traditional research methods. Allport, whose classic work on personal documents was earlier cited, also describes how data from a single case can be analyzed, placing emphasis on what he terms the "morphogenic" instead of the nomothetic. (2, pp 405-422) It is my hunch that such approaches as he describes are ideally suited to the study of art.

With persons beyond the child level, first-person singular statements about feelings, evaluations, intentions, images, technical problems, etc., are essential to an understanding of what the artist is doing. To elicit such statements requires, in my experience, a climate much like that described by Rogers (18) for psychotherapy, where the genuineness (congruence between statement and feeling) and empathy of the therapist, along with his unconditional positive regard for the client, are prime requisites. As in psychotherapy, to enter fully into this privileged arena is something of a trust and also an encounter not to be taken lightly. Any desire to manipulate the other seems out of place.

In my own work, I have interacted with the artist only between drawings. Usually the "inquiry" takes place before a drawing session, using in-process photos of the previous session's drawing as a basis for stimulated recall and exploration. The photographic in-process material, the finished works, the notes I and my assistant make each session, and transcriptions of the inquiries add together to form a rich picture of a person guiding and transcending himself in art. In addition, without using the verb "to teach" in a transitive, pre-thought sense, I cannot escape the feeling that the method described is a good one for learning in art. Certainly the artist's self-identity and motivation are increased thereby. I have insisted that these case histories constitute a kind of abstraction and "knowledge" which is a valuable addition to art education. Assuredly they keep one humble and a believer in a kind of humanism rather rare in today's educational world. They suggest that the workaday ego is a small and often misleading representation of the potentialities of any given person.

Identifying and Stating a Researchable Problem

Dale B. Harris

Many, perhaps the most, researchable problems come from questions we raise during the normal course of our teaching or our work with teachers in supervision or in our discussions with colleagues and with students. It is perhaps the rare problem that one identifies by sitting down and thinking about his area of work until he generates an "aha!" experience, locating the significant researchable issue. More often, a problem may occur in the course of professional reading, in reading professional discussions or controversies, or studying reviews of research in particular areas. As one studies research, frequently questions arise concerning the unanswered questions of another's research. Other problems are identified from review of the literature, locating areas of work which have been infrequently investigated. In this latter case the rational or "sitting down and thinking it through" approach may be quite appropriate.

Formulating a Problem Statement

Formally stated, the process of identifying and stating a researchable problem goes through the following stages or steps:

1. Stating the general question as it initially presented itself.
2. Clarifying this question; defining terms and making it as specific as possible.
3. If the question is broad, developing a number of specific questions that, if answered, would contribute toward answering the central question.
4. Distinguishing among the issues those which are descriptive, those which are normative, and those which are experimental in character.
5. Deciding what kinds of observations will provide the information necessary for answering (or beginning to answer) the question.

A Sample Question

From among questions raised in research workshops, two may be selected as illustrative. One of these as originally stated was, "What is the nature of the 'floundering' some beginning painting students go through? Do certain psychological conditions, such as anxiety,

confusion, hostility, fear, or rejection exhibit themselves to a certain degree in those students who identify themselves as flounderers?" A second question was, "How can we 'free up' those children whose drawings and paintings are too constricted?"

Restating the Initial Question

Taking the first question, one must break it down by asking additional questions or restating and dividing it into separate issues and then deciding the order in which one will seek the answers. For example, the first question raised above, calls up a number of further questions, such as, what are the behavioral indicators that the investigator believes will indicate floundering? Do students themselves know the experience of floundering? If so, are those who so identify themselves the same students that the instructor would name as "flounderers?"

Implicit Assumptions

It is well at this point to recognize that the question implies several assumptions. One certainly is that the students are mature, articulate, and well motivated individuals who are not simply "goofing off" because they dislike painting instruction. Clearly, and particularly if the art students were school children, we would have to separate the well motivated from the poorly motivated children whose indifferent approach to their work would confuse the issue. Secondly, such a question assumes that the investigator has had sufficient experience to identify the behavior of "floundering" and already know some of the overt signs of floundering. That is, one assumes that the investigator has a body of experience which can direct his initial observations; he will not observe all the behavior in the painting class--merely those aspects which relate to "floundering."

Formulating Specific Substudies

It is also necessary to recognize that the question very quickly breaks down into a number of distinct parts. One study would ascertain whether students themselves are aware of the experience of floundering, and whether these students are the ones the teacher would identify from his observation. Another would be a survey comparing two types of students. The behaviors of flounderers and non-flounderers would be compared with respect to such psychological characteristics as anxiety, confusion, rejection, hostility, fear, and the like. At this point it would be well for the investigator to talk with one of his psychologist friends, to identify acceptable behavioral indicators of the above nouns designating emotional states. This step is the necessary one of definition, and the selection of criteria. One must decide whether he will look for overt signs of these emotional

conditions in classroom behavior directly accessible to the teacher or will use psychological tests to get at dispositional traits. The psychologist may suggest certain personality tests which are sufficiently valid to investigate selected dispositional traits or characteristics. Here one must be cautious as, despite a great amount of research, there is still much confusion and debate in psychology as to the nature of dispositional traits.

Returning to the broader question as initially stated, it is clear that the person who raised this question wants to know whether perhaps something other than emotional conditions contribute to the floundering. Is some of this floundering in the nature of trial and error exploration rather than indicative of dispositional qualities or source traits? Clearly one aspect of the study should involve introspection and self report by students. Thus, investigation should focus on the process of learning, as well as on characteristics of the learner.

In any event, the investigator needs to spell out rather carefully what he means by floundering--the specific behavioral characteristics that he would identify as indicating this state in the learning of artistic skill. What about the following--asking the instructor many questions; making many false starts, and using many sheets of paper; wandering about the classroom; looking at the work of other students; expressing dissatisfaction with the media; offering a variety of excuses for not getting to work and staying at work--are all these signs of floundering? What other behaviors could signify floundering behavior?

A Checklist of Observable Behaviors

First, then, the investigator must identify in behavioral terms the condition of "floundering" as he defines it. Probably he will want to discuss these signs with colleagues to make certain that he has described a condition which other teachers also identify. Being able to state a commonly recognized situation gives some assurance that an individual is working with behaviors that have meaning and significance, and is not just expressing his own private concepts or peculiar interests.

Having satisfied himself that he has defined a meaningful cluster of criterion behaviors, the investigator next will want to make a list of the specific actions that students exhibit when they "flounder." He will also want to list a wide variety of illustrative verbal statements that students may make while "floundering." To provide these he will either draw on his store of accumulated observations or he will watch his class with an open eye and mind for a few days, making notes on the kinds of behaviors and statements which would lead him to say that students are floundering. Having listed such behaviors and statements

he would next ditto a preliminary checklist, arranging the items in convenient and logical groups. He would try out this checklist by observing for five or ten minutes each, several students whom he would identify as "typical flounderers," making all the behaviors that he observes and adding any which he had not included on the checklist. He would also wish to spend equivalent time with some typical non-flounderer to see that in fact relatively few if any of the behaviors on this checklist can be noted.

Having obtained a workable checklist, he would then proceed to systematic observation of all students in a given class. This task might require several periods of class time in which he would observe each student in turn for a short, specified time interval, checking off all the specific items of behavior which occurred in that interval of time for that student. He would go through all students in turn using equivalent amounts of time for observation for each student. On the basis of this simple checklist he could probably identify a stated number of students who show the most behaviors, and a similar number who showed the least behaviors in the time interval. These groups would constitute his contrasted groups of flounderers and non-flounderers.

His next step might be to interview these students to get their subjective experiences to substantiate his observations, or he might proceed to give psychological tests selected with the advice of a colleague, as previously noted.

It is well to recognize that in the above description we have involved several specific research tasks. We have defined terms, identified behaviors and verbal statements, conducted a simple survey which has certain normative characteristics to it. That is, the investigator has observed a group of students, each student for a similar period of time under similar conditions, and checked off a list of behaviors noted during that interval. He could make "normative statements" about the typical number of floundering behaviors that are likely to be exhibited in a short period of time in a class for beginning painting. He would have established a basis for further surveys of contrasted groups. He would also have a basis for identifying a group for an experiment to test which techniques might reduce the amount of floundering behavior. Such an experiment would require two groups of students equivalent in initial floundering behavior, to be treated, one with special teaching techniques designed to reduce floundering and the other, the control group, in which no special techniques are used.

A Second Example

Take the second question. "How can I free up children's paintings and drawings?" As stated, this is a methodological question, not a substantive research issue. Certainly it is not a problem you can make

observations on very readily. You wish to do something to achieve an objective, the objective being to make children's art work less stereotyped, more free and open. Again one starts with an assumption--that freedom in art expression is a good thing. This is a value statement. Indeed, there is a very respectable body of theory in art education that holds that freedom is a desirable thing, and there's much discussion in the professional literature concerning how this objective can be achieved.

Determining Criteria

For the descriptive portion of the research, the problem is to identify those aspects of drawings that are free vs. those that are stereotyped and constricted. So immediately we need criteria. Where does one get criteria? One can, of course, draw on theory. What kind of evidence would you look for in children's graphic art work that would indicate whether it's constricted and stereotyped or free and open? Someone may say--"The use of space and the nature of line." This statement is insufficient. Specifically, what usages of space and line suggest inhibition or freedon? I suppose most of us would agree that small figures, placed sparingly on a sheet of paper, or perhaps restricted to one small area of the sheet would suggest constriction. Lines made slowly and carefully, or perhaps lightly, uncertainly, or tentatively would also suggest "criteria"; already we are approaching verbal descriptions which should enable us to tell, when we look at a drawing, whether or not these criteria apply.

One might also consider that "movement" in the drawing has something to do with openness or constriction. Indeed drawings have been evaluated quite successfully according to the general criterion: "movement depicted or implied." one might say, "It's the number of straight lines vs. the number of curved lines." One might use such a criterion. Suppose, for example, one finds that the general rating of movement in the drawing is actually highly correlated with the proportion of curved to straight lines. Then he might conclude that these two criteria are really only one. But such a fact must be determined empirically; if you did find it, you'd need to use only one of those criteria, rather than both.

One might also look at themes--conventional vs. original themes in the content of art work. The notion is that conventional themes represent inhibition or constriction, and that unusual or "different" themes represent a form of freedom. In this we have another potential criterion, based on the idea expressed by the content. If it correlates only modestly with our other hypothesized criteria, we would have a basis for continuing to use it.

Reliability

Having established, from theory, some possible criteria, we must ascertain that various judges can observe them similarly. Our judgments of criteria must satisfy the public nature of science; they must show agreement, or be reliable. Therefore we must test out our criteria by having two or more judges independently classify a sample of drawings or paintings by these criteria, to establish that their judgments are in fact similar.

A Possible Experiment

What we've done is establish some criteria by which we can judge work to be constricted or free. However, we still have the question--how does one free up children's drawings? This now becomes a method question, suitable for an experiment. Many art teachers have rather definite ideas about how such may be accomplished. The investigator will try one of the methods suggested for "loosening up" children's work. He proposes to introduce a particular stimulus situation, to exercise a certain kind of instructional control, in order to produce in students' work a greater amount of the indices that he has identified as denoting freedom.

The experiment is science's most powerful technique, and it leads to the most useful and significant knowledge. Experimentalists often disdain descriptive research, but unless one has good description, he can't experiment, because he doesn't know what he is working with. The experiment comes relatively late in scientific development. It may well be that in the field of art education there are some areas that are ready for experiment, but there are surely many areas in which much more has to be known descriptively before we can identify what we want to do experimentally. Educators in art should not feel backward about simple descriptive work and hasten into experimenal work. There is plenty to be done in order to become familiar with the dimension of problems at a descriptive level.

We have said earlier that observation is only one part of research design and concern. In observational studies we ask "What happens?" and we can give descriptive accounts; we can make continued observations or limited observations; we can select critical instances; we can make process accounts, etc. Surveys answer the question "How are variables distributed in a population?"; and for surveys, as we indicated previously one uses interviews, questionnaires, schedules of items, or questions. The research problem of the survey is less that of accurate observation and description and more that of the sample--sampling subjects in such a way that one can make general statements from the sample to some theoretical larger population.

Correlational Studies

With instruments constructed for survey research--interviews, questionnaires, and schedules of various kinds (and remember that so-called personality tests are not tests of skills; they are really questionnaires or interviews)--one opens up a third area of work: Correlational studies. The research question now is "How are variables associated with one another?" One may work with simple relationships, or he may work with multivariate relationships--a number of variables intercorrelated. One may extend this procedure into factor analysis to examine, in a complex of intercorrelated variables, how many principal independent components or dimensions make up the complex of interrelated factors. Or one may work with prediction--by correlating variables (from which one hopes to predict) with particular outcome variables. Here, the problem of criterion measures is especially crucial, and we move immediately into the technicalities of measurement.

An Introduction To Measurement

by William Rabinowitz

To understand a person you must first be able to describe him and his behavior. This is equally true of teachers and children. To understand what takes place in the classroom it is necessary to describe it. This is also true of the products of behavior such as drawings and paintings. To understand such artistic products you have to be able to describe them.

We have been discussing various procedures for observing and describing the person, his behavior, and the products of his behavior. It should be clear that a description ideally ought to be complete and accurate. We have seen, however, that any description of a person is necessarily limited to a few of the many aspects of his behavior that are observable. Description is therefore selective.

In part, the selectivity of description involves a decision by the observer concerning what to observe; in part, it involves a decision about how to observe it. Naturally, we want to observe those aspects of an individual's behavior that are especially relevant to our purposes. Our description of these relevant aspects may be relatively unstructured, impressionistic, and qualitative, or it may be relatively structured, systematic, and quantitative. For many reasons, descriptions tend to be more useful as research data when they are, or can be, expressed in quantitative terms. To the extent that we succeed in making our descriptions quantitative and precise, we become involved in measurement.

Most people, when they think of measurement, think only in terms of paper-and-pencil tests. Although, such tests are an important facet of measurement in education, measurement implies far more than testing. In conducting descriptive research, a basic requirement is the development of adequate procedures for observing and describing behavior. Such descriptions of behavior may be considered measurements--sometimes crude and always imperfect, to be sure, but measurements nevertheless. It seems appropriate therefore to consider, some fundamental concepts related to measurement.

Fundamental Concepts of Measurement

An initial requirement in all measurement is defining the attribute or

property to be measured. If a property or attribute can be defined unambiguously, it can be measured--at least in principle. Notice that we do not directly measure a person or object. Instead, we measure certain properties or attributes of persons or objects. We measure the weight of a box, the height of a building, or the area of a field. Or we measure the intelligence of a child, the emotional climate of a classroom, or the creativity in a work of art.

In many instances, particularly when we are dealing with simple physical attributes, the meaning of the terms we use appears self-evident. We don't typically see any need to define what we mean by weight, height, or area. But when we want to study attributes that are of educational or psychological interest, we encounter definitional problems almost immediately. Psychologists - even after years of study - cannot agree on definitions for intelligence, emotional climate, or creativity.

You may have noticed that when you try to isolate and describe some interesting characteristics of drawings or paintings, the task presents certain problems. Many of these problems are directly related to the terms you use to describe the drawings - terms such as balance, tension, texture, rhythm, etc. The referents for these terms are not obvious. The attributes of the drawings to which these terms refer are not immediately apparent. Often you may find yourself challenged by your colleagues to define your terms. And often your definitions are not accepted by other members in your group. To the extent that you cannot specify clearly what you mean by the attribute-terms you use, it is a fundamental stumbling block in your efforts to do descriptive research.

I said earlier that descriptions are most likely to provide scientifically useful data when they are based on systematic observational procedures and are expressed in reasonably precise and quantitative form. Insofar as we are able to describe persons, objects, or events in this way, we are engaged in measurement.

Scientists have somewhat different conceptions of what measurement is, but among all of these varying conceptions there is a common idea. Measurement is the assignment of numbers to persons, objects, or events in accordance with certain rules. Whenever you have a rule- or a set of rules - that you can employ in assigning numbers to the phenomena you are observing, you are measuring.

The rules scientists use to measure vary, and this variation leads to the important concept of scales of measurement. Consider for the moment some of the familiar properties of numbers. Numbers can express equivalence or difference. Thus, 1 is the same as 1, but

different from 2. Numbers can express order of magnitude. Thus, 1 is less than (or comes before) 2, and 2 in turn is less than (or comes before) 3. Numbers can express equality of differences. Thus, 80 is as much greater than 70 as 70 is greater than 60. Finally, numbers can express equality of ratios. Thus, 80 is twice as large as 40 and 40 is twice as large as 20.

Each of these conventional properities of our numbe: system has a parallel in the scales of measurement we are going to discuss. And the exact nature of this parallel depends upon the rules we employ when we assign a number to a person, object, or event we are trying to measure.

Nominal Scales

Consider the simplest situation. As an example, we assign a group of persons to two categories, "Male" and "Female" and we call the males "0" and the females "1". The rule for assigning numbers here is clearly a very simple one. Each person can be unambiguously assigned to either the "0" group or the "1" group depending upon a particular property he (or she) possesses - namely his (or her) sex. All of the 0's are equivalent to one another with regard to this particular attribute, and all of the 1's are also equivalent to one another with regard to this particular attribute. But none of the 0's are equivalent to any of the 1's with regard to this particular attribute.

I could give many more examples of this type of measurement. We might, for instance, classify paintings into categories such as "oils," "watercolors," and "pastels" and use the numbers "1", "2", and "3"--or any other three different numbers--to designate the categories. Or we might classify schools into the categories, "Public", "Private-Parochial", and "Private-Nonparochial." Once again we could use three different numbers to designate the categories. In each of these examples it is easy to recognize the attribute we are trying to measure, and the rule we would use in assigning numbers to the objects--paintings or schools--being measured.

It may have occurred to you that this process doesn't really resemble measurement as we usually think of it. If this is your reaction, you are in distinguished company. Many scientists are also unprepared to call this use of numbers measurement. However, it is frequently referred to as a nominal scale, and I think it is useful to discuss it, although whether it should be considered measurement is debatable.

The numbers in a nominal scale simply indicate that two objects are equivalent or not equivalent in some respect--and nothing more. The distinguishing feature of a nominal scale, as you have probably

noticed, is the use of numbers as substitutes for verbal labels. Numbers used this way must be interpreted very cautiously. For example, the numbers that appear on the backs of football players constitute a nominal scale. We can be sure that "25" is a particular player with a particular name. In the same way, "50" is another, different player with a different name. The two men are different and they have different numbers as they should on a nominal scale. But notice that we can say nothing about the two men--other than that they are different--by examining the order of the two numbers of the difference between them.

Ordinal Scales

Often we can recognize not only that objects are equivalent or not equivalent to one another; we can also recognize that they can be ordered or ranked in some way. Thus we may observe five different art lessons and feel that we can order them on a characteristic such as, "Structure." We would assign a rank of "1" to the lesson that was most structured, a rank of "2" to the lesson that was next highest in structure, and so forth. This procedure for assigning numbers involves a different rule from the one that defined a nominal scale. Here we can discriminate order, not only difference. In effect, we are saying the event to which I assign a "1" is not only different from--it is also higher than--the event to which I assign a "2". A scale produced in this way is called an ordinal scale.

Ordinal scales are very common in every day life. When a movie reviewer describes a picture as a "3-star" picture and another as a "4-star" picture, he is using an ordinal scale. His use of the scale implies that any movie called a "4-star" picture is better than any picture called "3-star." Thus, the movies are being ranked or ordered.

Ordinal scales are also very common in educational and psychological work. The conventional letter grades, "A", "B", "C", "D", and "F" with which we assess academic performance constitute an ordinal scale. The scale implies that any student who receives an "A" in a particular course is a better student in that course than any student who receives a "B" and so forth. Though we use letters in assigning grades, we could as easily use numbers such as "4", "3", "2", "1", and "0" -- and this, of course, is exactly what we do when we compute a grade-point average.

Probably the most common examples of ordinal scales in educational and psychological work are the ubiquitous rating scales with which we try to measure everything from teacher effectiveness to pupil personality. Usually these scales have five steps--sometimes one or two more or less--with all of the steps decribed so that they fall into a

clear order. Thus, I could develop a scale to rate teacher-pupil rapport in the classroom which might look something like this:

Teacher-Pupil Rapport

1. Teacher and pupils work together in complete harmony; there is no evidence of any tension.
2. Teacher and puils work together in good harmony; there is little evidence of any tension.
3. Teacher and pupils work together in fair harmony; there is some evidence of tension.
4. Teacher and pupils do not work together in harmony; there is much evidence of tension.
5. Teacher and pupils are in obvious disharmony; there is evidence of almost constant tension.

This scale would clearly be an ordinal scale. You should recognize that none of the distinctions among the five steps in this rating scale indicates how much better the rapport is in one class than another. You can say that in a class rated "2" the rapport is better (or should be if the rating is accurate) than in a class rated "3". But you cannot say that it is as much better as a class rated "3" is better than a class rated "4".

Interval Scales

For some problems you must be able to show that equal differences in the magnitude of the events being measured are associated with equal intervals between the numbers assigned to these events. Thus, you must show, for example, that if you measure three events and they are assigned the numbers "15", "10", and "5" then the difference between 15" and "10" is the same as the difference between "10" and "5". In effect, you must be able to show that a 5-point difference represents the same thing everywhere on the scale.

When this condition prevails we are using numbers to indicate more than order. If we have demonstrated equality of intervals, we call the measurement procedure and interval scale. A familiar everyday example of such a scale is the Fahrenheit scale for measuring temperature. The difference between 20 F and 30 F is the same difference in temperature as that between 40 F and 50 F or 60 F and 70 F.

There are not too many characteristics of educational or psychological interest that we typically measure using interval scales. Consider intelligence as an example. It is frequently assumed that the IQ scale is an interval scale, but is that so? Is the difference between an IQ of 90 and 100 the same difference in intelligence as that between 60

and 70 or 140 and 150? Does a difference of 10 points have the same meaning, as a difference in intelligence, at all points on the scale? We often assume that it does, but that is only an assumption. Even when intelligence test data are converted to normalized standard scores, it is an assumption--not a demonstrated fact--that equal differences in standard score IQ's correspond to equal differences in intelligence.

Recall the discussion by Dr. Harris of the procedure we used to study the drawings of prospective teachers who were asked to "draw a picture of a teacher with a class."(16) We developed scales to measure in the drawings such characteristics as "Relative Teacher Emphasis," "Psychological Distance," "Teacher Initiative," "Traditionalism in Classroom Arrangement," and "Artistic Quality." Each of these scales is defined by a set of five sample drawings numbered from "0" to "4". In selecting the sample drawings we went to a great deal of trouble to select drawings that were equally distant from one another on the attributes being measured. We did this because we were interested in measuring the five attributes on interval scales. We could, of course, have developed ordinal scales with very little difficulty. We made an additional effort to develop interval scales because such scales, when they can be developed, permit more percise measurement and the use of more powerful statistical methods in data analysis.

Ratio Scales

There is a final type of measurement scale called the ratio scale. Certain events occur which match the number system not only in terms of increasing order and equivalence of intervals, but also in terms of the existence of a meaningful zero point. When a meaningful zero point exists in some attribute--we often say a true zero point--then it is possible to discuss ratios meaningfully. It is possible to say that A is twice, or one-half of, B on the attribute being measured.

The most familiar examples of ratio scales are the scales we use to measure such physical properties as length and weight. A string 5 inches long is exactly one-half as long as one that is 10 inches long. A box that weighs 20 pounds is twice as heavy as one that weights 10 pounds. We can make these statements meaningfully because in each case a unique and meaningful zero point exists. If we return to our earlier example, of the Fahrenheit temperature scale, you will note the importance of the requirement of an unique or true zero point on the scale. The Fahrenheit scale has a zero point; it happens to coincide with the temperature at which water freezes. It is not a true or unique zero, since there are clearly temperatures far below this point--represented as negative values. And for this reason it is not a ratio scale.

Ratio scales are rarely achieved in educational or psychological measurement. Consider intelligence once again as an example. Is there any sense in which we can say that a person with an IQ of 150 is twice as intelligent as one with an IQ of 75. We can, of course, say that two boxes, each of which weighs 75 pounds, weigh together as much as one box that weighs 150 pounds. But clearly two individuals with IQ's of 75 each cannot in any meaningful way be shown to be equivalent to one individual with an IQ of 150. The IQ scale may be an interval scale, but it surely is not a ratio scale. There is no unique and meaningful zero point in intelligence.

This, of course, is almost certain to apply to our efforts to measure characteristics of major concern to art educators. A meaningful zero point in the measurement of creativity, for example, is, I think, impossible to conceive. To the extent that I am right, it will not be possible to develop a ratio scale to measure creativity. And a statement such as "A is twice as creative as B" will be meaningless.

Let me try now to summarize very briefly what I have been saying about scales of measurement. We defined measurement as the assignment of numbers of persons, objects, or events in accordance with certain rules. Depending upon the rules employed, we may measure using a nominal, ordinal, interval, or ratio scale. In most of the situations in which you as art educators are likely to be measuring, you will probably be using nominal or ordinal scales. In some cases you may be working with interval scales, but that is not likely to be very frequent. Except when you measure a painting's size or the weight of a piece of sculpture, you will not have many occasions to employ ratio scales in your descriptive research.

Errors in Measurement

Let us assume now that you have a measurement procedure. Perhaps you are measuring aspects of behavior in the classroom or some important characteristics of children's art products. One of the problems you encounter almost immediately is the measurements are not completely accurate or dependable; they are subject to what is called error. We have to spend some time discussing error, because the presence of error in measurement means that scores are not, and cannot be, completely reliable.

We begin then with the recognition that no measurement is perfect; all measurement contains error. And this is true of <u>all</u> measurements -- in the physical sciences as well as the behavioral sciences. Fortunately for physical scientists, the magnitudes that they are usually dealing with are very large relative to the errors of measurement. Since the error sizes are typically relatively small and the

magnitudes they are trying to discriminate are relatively large, the physical scientist can usually measure with a great deal of confidence. In psychology and education, the situation is very different. We are often trying to measure small differences in some attribute, but the measurement errors are relatively large. As a consequence when we detect a difference in our measurements we often cannot be sure if it reflects a true difference between the objects being measured or is due to error.

In developing a measurement procedure, one of the first and most important tasks you face is reducing the size of the errors of measurement as much as possible. To the extent that errors are very prominent in your measurements, they will not -- and cannot -- relate to anything. Research based on measurement procedures that are unreliable -- that is, subject to large errors -- is doomed to fail. So you must try to keep error of measurement as small as possible. To do this you must understand what error is and how it gets into measurements.

Systematic Errors

As an aid to this discussion, I am going to make a distinction between systematic and random errors. Let me discuss systematic errors first. Systematic errors occur whenever you are measuring, and the numbers you assign to certain persons, objects, or events are systematically either too high or too low. This could happen for many reasons and in many different ways. Suppose I am trying to measure the intelligence of everyone in this room. I do not have an intelligence test; all I have as raw data are my observations of your behavior. So probably I would have to rate your intelligence on the basis of these observations. Let's say that I had a rating scale, an ordinal scale with five steps. Suppose further that the nature of my measuring process is such that I tend to think of men as more intelligent than women. For the men, I tend to resolve all of my doubts in favor of assigning higher ratings; for the women, I tend to resolve all of my doubts in favor of lower ratings. That is a systematic tendency on my part to rate men too high and women too low. On the perfectly reasonable assumption that intelligence is distributed equally between the sexes, any tendency on my part to judge men higher than women is a systematic error. The effect of this error on my measurements is highly predictable once you know the nature of my bias.

Suppose, as a further example, you wanted to measure the extent to which a group of teachers were able to establish rapport with pupils. You might ask the principals of these teachers to rate them on this characteristic. It is highly likely that the principals' ratings would not be pure measures of rapport. The principals, being human, would

very likely give higher ratings to teachers whom they found highly cooperative and lower ratings to teachers whom they found highly uncooperative. This, of course, is a systematic error. The teachers' scores on rapport are being systematically elevated and depressed by a factor that is unrelated to what you are trying to measure.

In developing a measurement procedure it is important to anticipate the systematic errors that are likely to be operative so that their influence can be reduced. For example, suppose you were doing a study involving observations of teachers in the classroom. If one of your observers tended to rate the pretty teachers higher than the not-so-pretty teacher; he would be introducing a systematic error into your measurements, on the assumption that you were not trying to measure prettiness. Once you discovered this, you would have to do something about it. You would either re-train this observer or drop him from your study.

To pursue this example further, suppose the bias of this observer was very strong. Let's say that every time he observes a pretty teacher he rates her "5" and every time he observes a homely teacher he rates her "1". If this were the case, you could never discover anything about the classroom behavior of these teachers, because their scores were totally determined by the observer's bias. The systematic error is so large in this case that it completely overwhelms the property you are trying to measure. This, of course, rarely takes place, and when it does it is very easy to detect. What is more common is for much smaller error tendencies to operate. Though smaller, such error tendencies can nevertheless be large enough to reduce markedly the accuracy of your measurements.

Another example of systematic error is something that occurs in a quite different context, and that is when people report -- through questionnaires or interviews -- their attitudes or feelings. In such situations, it is common for individuals to distort their responses in order to create a favorable impression. For example, suppose you were giving a personality inventory to a group of individuals, all of whom were applying for a job as a salesman for a life insurance company, and you had questions on the inventory like: "Do you enjoy talking to people?" and "Do you feel comfortable when you're trying to influence somebody to make a decision?" Now consider an applicant for a position as life insurance salesman. How will he respond to questions like these? He will almost certainly answer "yes," whether or not he feels that way. His scores are therefore going to be systematically distorted in the direction of appearing more like a life insurance salesman than would probably be the case if he were not taking the test for that purpose.

The same thing can happen in a research study with teachers. If you

give teachers personality or attitude inventories to fill out, they are almost always going to try to put their best foot forward. We worked some years ago with a test called the Minnesota Teacher Attitude Inventory. It's a test that has statements like, "Most children are obedient." The response options are "Strongly Agree," "Agree," "Uncertain," "Disagree," and "Strongly Disagree." There are 150 statements with five possible responses to each statement. In one study we gave this test to prospective teachers and then collected their test papers. Then we gave the test once again, but the second time we said, "Now take it and try to get the highest possible score." When we compared the scores on the second testing with the first, we didn't find that there was much difference. As a matter of fact, there was almost no difference whatsoever. This makes you feel that the first time the prospective teachers took the test, they were trying to get the highest possible score. In effect, they were faking a good performance -- distorting their "true" attitudes in order to earn as high a score as they could. This, of course, is a type of systematic error.

Suppose you want to observe and measure certain aspects of the typical classroom behavior of teachers. You might make it a practice to advise teachers well in advance of your visit of your desire to observe them at a particular time. This seems like an appropriately courteous action on your part, but it is likely to mean that the behavior you observe will, to some unknown extent, not be typical. It will almost certainly be systematically modified by the teacher in order to create a good impression. The teacher exhibits her best behavior which may or may not be typical of her.

I noted earlier that rating scales are widely employed as measurement devices in educational and psychological work. Though very popular ratings do not typically yield very satisfactory measurements. Part of the difficulty is that they are subject to at least two types os systematic errors.

One of these is generosity error, that is, the tendency of raters to give ratings that are too favorable. Principals tend to rate their teachers either "Excellent" of "Very Good." Only the most troublesome or inadequate receive low ratings. Such ratings are of little value since they do not discriminate effectively among the individuals being rated. In general, generosity error is likely to be very prominent whenever supervisors rate subordinates. And this suggests, in part, why ratings are too generous. The supervisor is likely to feel a greater loyalty to those with whom he works -- no matter how inadequate they may be -- than he does to those who are requiring him to make the ratings.

The other type of systematic error in ratings is called the halo effect.

To understand this type of error you must recognize that usually when ratings are employed, the rater is expected to rate individuals on several separate traits. The person who wants the ratings hopes to get several, relatively independent pieces of information about each ratee. And that is the reason for the use of several rating scales. In practice, however, raters tend to be strongly influenced in assigning their ratings by the general, overall impression they have of each man. Thus, the rater does not usually make the careful discriminations implied by the necessity to rate a number of specific traits. Instead, the ratings of these specific traits tend to reflect the rater's general opinion of the individual's merit or worth. The halo effect means that ratings of specific traits are underdifferentiated -- or, to put it another way, too highly correlated with one another. In any event, the effect of this error is systematically to obscure the pattern of traits within the individuals being rated.

Random Errors

I could continue discussing sources of systematic error, but it seems appropriate to consider briefly the other type of error -- random error. You can recognize the difference between what I am calling systematic and random error by noting that systematic error produces a predictable effect whereas random error does not. The biases that are lumped together and referred to as systematic error have a predictable tendency to elevate scores in some cases and depress them in other cases. Generosity error, for example, tends to produce scores that are too high.

Random error is inherently unpredictable. It occurs because of a myriad number of influences most of which are so poorly understood that we often say that random error is due to chance factors. We invoke chance to "explain" random errors of measurement because in a fundamental sense it really cannot be explained.

It is important to recognize, however, that random errors will have unpredictable effects, sometimes elevating, sometimes depressing scores. Thus an obtained score, derived from some measurement procedure, is distorted in some unknown way by random errors of measurement which cause it -- in an inherently unpredictable manner -- to be either too high or too low.

We can illustrate some of these rather abstract concepts by reference to a hypothetical example. Suppose I am interested in measuring the effectiveness with which a group of art teachers conduct their art classes. As part of this effort, I visit Miss Suzy Jones on April 20 at 9 A.M. for 30 minutes. At the end of my 30-minute visit, I give Miss Jones a rating and this rating represents her score as far as my

investigation is concerned. You can assume that the other art teachers are measured in the same way.

It is probably very clear to you that this method for measuring a teacher's effectiveness leaves a great deal to be desired. Our measurement of Miss Jones, and the other teachers, is derived from a single observation. We know nothing about how typically Miss Jones behaved on that particular occasion. Specifically, how can we be sure that Miss Jones was not atypically exciting or dull on that day? Since the observation was taken at 9 A.M. we may assume that Miss Jones is not a better, or worse, teacher in the morning than in the afternoon, but do we know this? And since the observation was terminated at the end of 30 minutes, we are probably assuming that this is an adequate sample of Miss Jones' behavior. But Miss Jones may be one of those teachers who gets better, or worse, as a lesson continues. Finally, how shall we interpret the use of a single observer to visit Miss Jones and the other art teachers? Is the definition of "effectiveness" in art teaching so unambiguous that competent observers will always agree on the score that each teacher should receive?

The questions I have been posing suggest the sources of random errors of measurement. Behavior is unstable; it changes from one situation or occasion to another. The measurement standards of different observers are not uniform; they vary from one observer to another. This means that a score based on a single observation by one observer is not very trustworthy. You cannot depend upon it to reflect the true status of the individual observed.

This concept of "true" status requires some examination. Let's go back to Miss Jones and the other art teachers. We were trying to measure their effectiveness, but what does that mean? Without trying to define "effectiveness" it should be clear that we are trying to measure a general property of each teacher's behavior. How any specific teacher functioned on any particular occasion - as observed by a particular observer - is of no great consequence, since we are interested in general behavioral properties, not specific events. Each teacher's "true" effectiveness could be thought of as an average of her behavior across all possible occasions as observed by all possible, competent observers. Clearly, to measure true effectiveness, conceived of in this way, is an impossibility. But how can we approximate it? It should be obvious that we need many observations by several observers.

This then is the principal way in which we reduce random errors of measurement -- by taking a larger number of observations. This is a highly generalizable principle; it applies to paper-and-pencil tests as well as observational measures. All other things being equal, a

longer test is more reliable than a shorter test. Adding test items reduces the relative contribution of random errors of measurement to individuals' scores. In the same way, a measure of an art teacher's effectiveness becomes more reliable as it is based on a larger number of observations.

This may be an appropriate point to conclude our discussion of measurement. I hope that it is clear that my comments were quite general, and deliberately so. Measurement can be a very technical area, and to go much beyond my very general remarks will almost surely involve technical concepts of increasing complexity. For example, we cannot explore such concepts as validity or reliability without some statistical understandings. The going gets rougher as you continue. I hope, however, that I have encouraged you to see measurement as comprehensive and interesting, and that you will continue beyond this point.

Postscript

The preceding pages have presented the major content that was the focus for the three days of each of the Research Training Institutes on Descriptive Research. The purpose of the publication is similar to that of the Institutes, that is, to inform and to encourage participants (readers) to get involved with research methods in their own situation.

With all good intentions, those who have little or no previous background in research methods, may still face, with some puzzlement, the question, "Where do I start?" At the risk of oversimplification, the following may be a way to begin. This is learning exercise, similar to one that was used on the first day of the Institutes.

1. Collect a small number of drawings from your class (or someone else's), at least 10 but not more than 20, to start with. Work done on the same size paper will be more convenient.
2. Lay them on the floor or pin them on the wall, in any arrangement where you can look over all the work with ease.
3. You are now ready to make your first observations. With pencil and paper in hand, look over all the drawings carefully and begin to list the attributes (characteristics) that are present in some or all of the drawings. For example you may note such attributes as: subject matter differences; some used shading, others did not; all had shapes but some had more than others, and so forth. A note of caution is important as you proceed. We are concerned at this point only with what we can observe. You may be tempted to include inferences, such as "Some students didn't finish their work." This, of course, is an inference you have made from something you have observed but is not the direct observation itself. As you continue, it is likely that you will find that the task becomes more involved than you had anticipated, as you move from simple to more complex phenomena. For example, in your first observations you may have noted that all the works had lines in them and that some had more lines than others. As you proceed you may decide that having few or many lines seems less important than the qualities of lines, and you will begin to note line characteristics such as straight or curved, light or dark, thick or thin.
4. After you have listed 5 or 6 attributes, it will be wise to test these by sorting the drawings according to each of the attributes. You may separate all of the drawings that have shading from those that do not. At this point you may feel disturbed that there

is so much variety in the pile of drawings with shading; some have very little compared to others. In trying to separate them according to the amount of shading -- you may find yourself thinking in terms of: none; a little bit; a lot; all shaded. In doing so you have developed an ordinal scale that can be used for measuring the amount of "shading" that is observable in the drawings.

5. As the number of attributes increases, you are likely to encounter a certain amount of overlapping of terms. For instance, you may find that a "light line" may refer to one that is "thin", one that is "light in value", or a combination of these. In order to be more specific you will need to define your terms in order to clarify their meaning. If you decide that a "light line" is a line that is light in value, your descriptions can become more precise since you can now refer to "light thin lines" or "light thick lines".
6. Eventually you will realize that the task seems unending and decide to stop. The exercise is in theory an endless one increasing in precision as you discover attributes in the fine detail of the drawings and define your terms. Having read this far, you may decide not to try this exercise at all, since you have just read about it. However, do not jump to this conclusion too quickly; reading about it is no substitute for the experience of doing it. This opportunity to test your own perceptions in discovering attributes, practice in clarifying meanings, and developing rating scales will pay dividends when you begin working on specific problems. The section by Dr. Harris on "Identifying a Researchable Problem" will help you take the next step and you are on your way toward getting involved with research methods.

Descriptive research is that category of research that attempts to describe "the way it is" with more precision and accuracy than one can obtain through casual observation. It is easy to say that each child's drawing is a unique expression, but by the recording and analyzing of careful systematic observations of children's drawings we can begin to sort out what similarities and differences exist. When we have done this, we can begin to verify and give more specific meaning to the assumption we have expressed. However, in comparison to past history, "the way it is" today may be only a temporary condition because of the rapid changes that are taking place throughout all facets of our society.

The need for ongoing descriptive research has never been more critical than it is at this time. Anyone who has been involved with public education for the past decade is well aware that the youth of today are quite different from those in school during the 1950's. But when we begin to discuss "how" they are different, we rely on imperfect memories and describe in vague generalities. Changes in art programs

and the development and inclusion of new art materials and new art forms is on record. Knowledge of changes in students and their effects on behavior in art remain speculative since we have little documentation in this respect to make comparisons.

No research of any kind will provide the answers to the "ought" questions that we face every day. What curriculum changes "ought" we to be making to keep up with times? What knowledge "ought" students to be learning in order to cope with the problems of the future? What criteria "ought" we to use to evaluate students? The answers do not lie in the findings of research studies, but a backlog of descriptive materials about students and student behavior in art can be analyzed for changes and developmental trends that can help provide a basis for making more intelligent decisions concerning these kinds of questions. Without more reliable information we are left to rely on the fads and fashions of educational innovation.

Art educators trained in research methods are relatively few, and their combined number can not provide the descriptive data that is needed. We can only hope that more interest and active involvement in gathering descriptive material will come from teachers and supervisors working in the schools and colleges.

BIBLIOGRAPHY

1. A Seminar in Art Education for Research and Curriculum Development; Cooperative Research Project No. V-002; Edward L. Mattil, Project Director, Pennsylvania State University, University Park, Pennsylvania, 1966.

2. Allport, Gordon W., "The Unique and the General in Psychological Science", Journal of Personality, 1962.

3. Allport, Gordon W., The Use of Personal Documents in Psychological Science New York: Social Science Research Council, Bulletin 49, 1942.

4. Art Education; edited by W. Reid Hastie, 64th Yearbook of the National Society for the Study of Education (Part II), 1965.

5. Barker, R. G., and Wright, H. F. Midwest and its Children, Evanston, Illinois, Row Peterson & Co., 1954.

6. Barker, R. G., Ecological Psychology, Stanford, California, Stanford University Press, 1968.

7. Eisner, Elliot W., "The Development of Information and Attitude Toward Art at the Secondary and College Levels", Studies in Art Education volume 8, Number 1, Autumn, 1966.

8. Eng, Helga, The Psychology of Children's Drawings. London: Routledge & Kegan Paul, 1931.

9. Eng, Helga, The Psychology of Child and Youth Drawing. New York; Humanities Press, 1957.

10. Exemplary Programs in Art Education; edited by Stanley Madeja, National Art Education Association, 1969.

11. Goodenough, Florence I, Measurement of Intelligence by Drawing, New York, Harcourt Brace, and World, 1926.

12. Harris, Dale B., Children's Drawings as Measures of Intellectual Maturity, New York: Harcourt Brace, 1963.

13. Hildreth, Gertrude, The Child Mind in Evolution: A Study of Developmental Sequence in Drawing, New York: King's Press, 1941.

14. LaPage, Geoffrey, Art and the Scientist, Bristol: John Wright and Sons, Ltd., 1961.

15. McGurdy, H. G., "Group and Individual Variability on the Goodenough Draw-A-Man Test", Journal of Education Psychology, 1947, 38, 428-36.

16. Rabinowitz, W. and Travers, R. M., "A Drawing Technique for Studying Certain Outcomes of Teacher Education" The Journal of Educational Psychology, Vol. 46, No. 5, May, 1955.

17. Report of the Commission on Art Education: edited by J. Hausman, National Art Education Association, 1965.

18. Rogers, Carl R.,"Learning to be", in Farber, Seymour, and Wilson, editors, Conflict and Creativity, New York, McGraw Hill, 1963.

19. Sax, Gilbert, Empirical Foundations of Educational Research, Engelwood Cliffs, New Jersey, Prentice-Hall, Inc., 1968.

20. Seminar on Elementary and Secondary School Education in the Visual Arts: Cooperative Research Project No. V-003; Howard Conant, Project Director, New York University, New York, N. Y. 1965.

APPENDIX

A. Selected References 79
B. Preconference Program 81
C. Institute Evaluation, William Rabinowitz..... 83
D. Pre and Post Test, with Key..... 89
E. Participants Institute Evaluation Form 99

APPENDIX A

Selected References Recommended for those Interested in Descriptive Research

Journals

American Journal of Art Therapy. (Formerly, The Bulletin of Art Therapy). (4 times a year). 6010 Broad Branch Rd., N.W., Washington, D.C. 20015. (Subscription $5.00 a year).

Journal of Aesthetic Education. (4 times a year). University of Illinois Press, 1002 W. Green St., Urbana, Illinois 61801. ($7.50 a year).

Journal of Aesthetics and Art Criticism. Published quarterly by the American Society for Aesthetics at Wayne State University: College of Liberal Arts and University Press, and The Cleveland Museum of Art, Mt. Royal and Guilford Avenues, Baltimore, Maryland 21202.

Journal of Creative Behavior. (4 times a year). Creative Education Foundation, State University College at Buffalo, 1300 Elmwood Ave., Buffalo, N. Y. 14222. ($8.00 a year).

Review of Educational Research. (Now published 5 times a year). American Research Education, 1126 16th St., N.W., Washington, D.C. 20036. This journal cyclically (3 year cycles) reviews fields and research topics. Last reference on Art Education: Vol. XXXVII, No. 2, 1967.

Studies in Art Education. A journal of issues and Research in Art Education. (3 times a year - Fall, Winter & Spring). 1201 16th St., N.W., Washington, D.C. 20036. ($5.00 a year, or as part of the comprehensive membership, along with one gift new publication, a $20.00 membership).

Books and Reviews

A Seminar in Art Education for Research and Curriculum Development. E. L. Mattil, ed. C.R.P. No V-002. University Park: The Pennsylvania State University; 1966.

Arts and Humanities Educational Research Projects. U. S. Department of Health, Education, and Welfare. Office of Education, Bureau of Research, Arts and Humanities Program, Washington, D.C., 20202; July 15, 1969.

Bloom, et.al. Taxonomy of Educational Objectives. Handbook I: Cognitive Domain (1956), Handbook II: Affective Domain (1964).

Crenbach, L. J. Essentials of Psychological Testing. Harper; 3rd. ed.; 1969.

Dissertation Abstracts. Ann Arbor, Michigan: University Microfilms. (Title varies: 1938-1951, Microfilm Abstracts; 1952 to present, Dissertation Abstracts).

Encyclopedia of Ed. Res. Robert L. Ebel, ed. New York: MacMillan; 1969. Also see: Encyclopedia of Ed. Res. Chester W. Harris, ed.; 1960.

Gage, N. L. Handbook of Research on Teaching. Chicago: Rand McNally; 1963. (Now being revised.)

Games, P. A. and Klaire, George R. Elementary Statistics. McGraw-Hill; 1967.

Harris, Dale B. Children's Drawings as Measures of Intellectual Maturity. Harcourt, Brace, World; 1963.

Kerlinger, Fred. Foundations of Behavioral Research. Holt, Rinehart and Winston; 1964.

Lanier, Vincent. Doctoral Research in Art Education. University of Southern California; 1962. (updated with later supplements).

APPENDIX B

PRECONFERENCE PROGRAM*

Research Training Institutes for
Descriptive Research

DAY ONE

8:00 - 9:00 AM	Registration
9:00 AM	FIRST SESSION - LARGE GROUP Welcome Introduction of Staff Program Orientation Pre-Test
10:30 - 12 Noon	SECOND SESSION - LARGE GROUP Lecture and Discussion (parts of 1 & 11) Drawing problem for participants
12 Noon - 1:30 PM	Lunch
1:30 - 3:00 PM	THIRD SESSION - SMALL GROUPS Using Drawings: Identify criteria, define critera, develop simple scales of variables.
3:00 - 5:50 PM	FOURTH SESSION - LARGE GROUPS Lecture and discussion (parts 1 & 11) Small Group reports on work Questions Evaluator's comments on events of Day One

* This is a resume of the programs with brief descriptions of the contents of each session. The information in the parentheses refers to the parts of this publication that were used in the lectures.

DAY TWO

9:00 - 9:45 AM	FIFTH SESSION - LARGE GROUP Lecture and Discussion (part of IV)
9:45 - 11:00 AM	SIXTH SESSION - SMALL GROUPS Work on problem selection and refining it
11:00 - 12:30	SEVENTH SESSION - LARGE GROUP Lecture and Discussion (Remainder of IV and part V)
12:30 - 2:00 PM	Lunch
2:00 - 4:00 PM	EIGHTH SESSION - SMALL GROUPS Continue refining researchable problem and begin development of measuring instruments where appropriate.
4:00 - 5:30 PM	NINTH SESSION - LARGE GROUP Lecture and Discussion (remainder of V) Discussion of problems encountered in small groups Questions Evaluator's comments on Day Two

DAY THREE

9:00 - 10:30 AM	TENTH SESSION - LARGE GROUP Lecture and Discussion (111)
10:00 - 12 Noon	ELEVENTH SESSION - LARGE GROUP Questions and discussion of problems
12:00 - 1:30 PM	Lunch
1:30 - 3:00 PM	TWELFTH SESSION - SMALL GROUPS Final work session. Preparing final form of research problem, design, and measuring instruments.
3:00 - 5:00 PM	THIRTEENTH SESSION - LARGE GROUP Post-Test Reports on work from small groups Final questions and discussion Evaluator's comments Farewell

APPENCIX C

EVALUATION REPORT

Prepared by William Rabinowitz

1970 NAEA

Preconference Educational Research Training Program

In Art Education

1. Evaluation Planning

The 1970 NAEA Preconference Educational Research Training Program consisted of our sessions as follows:

March 7-9, 1970/ Sheraton Hotel/ Philadelphia, Pennsylvania
(NAEA Eastern Regional Conference)

March 19-21, 1970/ Sheraton-Schroeder Hotel/ Milwaukee, Wisconsin (NAEA Western Regional Conference)

April 12-14, 1970/ John Marshall Hotel/ Richmond, Virginia
(NAEA Southeastern Regional Conference)

April 23-25, 1970/ Hilton Hotel/ Portland, Oregon
(NAEA Pacific Regional Conference)

Each of these sessions preceded the regularly scheduled regional meeting of the NAEA.

A decision concerning the U. S. Office of Education's willingness to sponsor the program was communicated shortly before the Philadelphia session was scheduled to begin. There was therefore little time available in which to plan a systematic and thorough evaluation. Nevertheless, it was possible for members of the planning committee (Drs. G. Kensler, D. B. Harris, E. Mattil, S. Madeja, and J. Mahlmann) to meet on February 27, 1970, and to plan details of the program including aspects of the evaluation.

It was agreed that the formal evaluation would consist of a multiple choice

objective test (to measure student comprehension of basic principles of descriptive research) and a questionnaire (to measure student reactions to various facets of the organization, presentation, and content of the program). Preparation of these instruments was to be the responsibility of the Program Evaluator, Dr. W. Rabinowitz.

In preparing the objective test, the Evaluator worked closely with the Research Specialist, Dr. D. B. Harris. Since Dr. Harris had primary responsibility for conducting the large-group lecture sessions during the program, it was important that the test cover the same material he planned to present. Although the "fit" between Dr. Harris' Instruction and the objective test was reasonably close, it was not possible in the limited time available to construct a test that perfectly reflected the concepts presented in the program.

In preparing the questionnaire, the Evaluator worked closely with the Project Director, Dr. G. Kensler, and Dr. J. Mahlmann of NAEA. Since Drs. Kensler and Mahlmann were primarily responsible for the planning and administration of the program, it was important that the questionnaire cover aspects of the program in which they were especially interested.

Copies of the objective test and questionnaire used as evaluation instruments are appended to this report.

11. Evaluation Activities

Each preconference program consisted of three days of instruction and discussion. During the first meeting of the participants, immediately following some brief words of welcome and introduction, the objective test was administered. All copies of the test and answer sheets were collected and scored at this time. The objective test was administered once again during the afternoon of the final day of the conference. Answer sheets were collected, but participants, at this time, were permitted to retain copies of the tests and were given an answer key. The questionnaires were also administered during the afternoon of the final day of the conference.

In addition to administering the test and questionnaire, the Evaluator functioned as an observer of the lecture and small-group sessions. During "breaks" the Evaluator discussed aspects of the program with staff and participants. At the end of each day, during a brief session, the Evaluator reported informally to staff and participants on a variety of matters of concern. These feedback sessions included, for example, discussions of the test results, areas of special concern among participants, similarities and differences in the activities of small groups, etc.

III. Evaluation Results

The evaluation results will be discussed separately for the objective test and the questionnaire.

Objective Test

The objective test was administered as both a pretest and post test, making it possible to assess directly the gain of each participant. The results at the four preconference sessions did not differ significantly from one another, and they have therefore been combined. Based on a total of 89 participants, for whom both pre- and post test data were available, the following results were obtained:

	Mean	Standard Deviation
Pretest	22.9	4.2
Post Test	27.3	3.6

The mean gain of +4.4 points is significant at the .01 level (t = 10.82, d.f. = 88).

Although the gain was statistically significant, it was clearly not very impressive. Almost all of the students showed some gain in score (a few gained more than 10 points), but many complained that the post test did not adequately reflect what they had learned during the preconference sessions. In some instances, this was attributed to the imperfect correspondence between the test and the content of the instruction; in other instances, it was attributed to difficulties with the multiple-choice format. It seems highly likely that, although the test provided evidence of significant, relevant learning, it did not reveal the full extent to which this learning occurred during the program.

Questionnaire

Completed questionnaires were available from 87 participants. As in the case of the objective test, the results at the four preconference sessions did not differ significantly from one another, and they have therefore been combined. The responses of the participants are summarized below for the eight questionnaire items in which a direct tabulation of response is possible.

1. How would you evaluate the organization of the program?

   Percent

   ____ Excellent organization in clear and meaningful sequence

Percent	
54	Generally well-organized
11	Adequately organized, but could have been better
0	Inadequately organized, too much confusion
0	Very disorganized, no apparent structure

2. How realistic and attainable were the objectives of the program?

Percent	
17	Very realistic and easily attainable
49	Most of the objectives were clearly attainable
22	Many of the objectives were attainable, but some were not
1	Most of the objectives were not attainable
1	The objectives of the program were clearly unrealistic
10	No response

3. To what extent did the program meet your prior expectations?

Percent	
21	The program exceeded my expectations
53	My expectations were generally well-met
13	The program was O.K., but not all my expectations were met
10	I expected more from the program than I obtained
0	The program completely failed to meet my expectations
3	No response

4. To what extent did the program deal with questions and problems that are applicable to your needs?

Percent	
34	The program was exceptionally well-related to my needs
34	The program was generally well-related to my needs
24	The program was adequate in terms of my needs
7	The program was only slightly related to my needs
0	The program was completely unrelated to my needs
1	No response

5. Was the length of the program appropriate?

Percent	
5	Considering what was covered, the program was too long
63	The length of the program was just right
30	Considering what was covered, the program was too short
2	No response

6. Was the difficulty level of the program appropriate?

Percent	
6	Considering my background, the program was too difficult
74	The difficulty level of the program was just right
16	Considering by background, the program was too easy
4	No response

7. In general, how would you evaluate the quality of the instruction?

Percent	
55	Outstanding
37	Good
6	Satisfactory
1	Inadequate
0	Poor
1	No Response

8. Would you recommend that the NAEA continue to sponsor research programs in the future?

Percent	
99	Yes
1	No

The questionnaire responses clearly indicate that, in general, the participants were quite positive about their experience in the program. More than 90 percent of the participants, for example, considered the quality of the instruction "excellent" or "good", and almost 70 percent felt that the program was well-related to their needs. The program was generally considered well-organized and appropriate in length and difficulty level. Almost two-thirds of the participants described the objectives of the program as clearly or easily attainable, and almost three-fourths felt that their expectations were well-met or exceeded by the program. Ninety-nine percent recommended that the NAEA continue to sponsor such programs in the future.

The questionnaire encouraged respondents to make any relevant comments concerning the program when they felt so inclined. An examination of these comments is instructive. For example, a number of constructive criticisms were offered. Several participants felt the need for a stronger orientation concerning the objectives and overall purpose of the program. Some suggested that the objectives should have been "spelled out" during the first session of the program; others felt that a more detailed and explicit description of the program should have been distributed in advance. The small group sessions were, in general, not evaluated as highly as the large group lectures. There was a feeling, expressed by some participants, that the purpose of the

small group sessions was unclear. Others were disturbed by what they considered a lack of relationship between the lectures and the small group activities. As one participant put it, "We should have had more structure and direction from the leaders in the small groups. It was hard to relate what we were doing to the concepts discussed in the lectures."

Although most of the comments were implicitly or explicitly critical, they did not significantly modify the general picture of satisfaction with the program. Most of the participants would probably have agreed with one who said, "My expectations were vague when I came, and I didn't really know what to expect. I've not been disappointed, however. This has been much better than previous seminars."

APPENDIX D

FORMAL EVALUATION TEST WITH KEY

1970 NAEA

Preconference Educational Research Training Program
In Art Education

Directions: Record your answer on the separate answer sheet. Do not quess wildly.

1. The first step in any research study is

   a. selecting the statistical procedures that will be used.
   b. choosing a good design.
   c. selecting the persons who will be studied.
   d. specifying the question(s) to be answered.

2. In order for a question or problem to provide a useful basis for research, it must be

   a. free from any implied value judgments.
   b. stated in quantitative, rather than qualitative, form.
   c. answerable through some empirical procedure.
   d. based on an explicit educational theory.

3. Which of the following is most likely to result from formulating the objectives of instruction in behavioral terms?

   a. Students will learn in a more efficient manner.
   b. Measures of student achievement based on instructional objectives will be developed.
   c. Subjective factors in the evaluation of the student will be eliminated.
   d. The student's behavior will be the basis for establishing the goals of instruction.

4. Characteristics of persons or things which can assume different values are called

   a. distributions.
   b. observations.
   c. data.
   d. variables.

5. An operational definition of a variable is one expressed in terms of

   a. mathematical symbols.
   b. action-related postulates.
   c. antecedent-consequent relationships.
   d. observable events.

6. Which of the following represents the crudest level of measurement?

   a. Joe can draw.
   b. Joe can draw very well.
   c. Joe can draw better than Bill, but not as well as Mary.
   d. Joe can draw better than he can paint.

7. No matter how a research study is conducted, the data ultimately consist of

   a. statistical indicators.
   b. reliable measurements.
   c. human observations.
   d. unwarranted inferences.

8. A product scale would be most appropriate to evaluate performance in

   a. drawing.
   b. arithmetic.
   c. spelling.
   d. reading.

9. Research studies which attempt to determine the relative incidence, distribution, and inter-relations among various sociological and educational variables in populations by selecting and studying samples are called.

   a. surveys.
   b. field experiments.
   c. normative studies.
   d. cross-sectional investigations.

10. Observations planned in advance are preferable to ratings based on past experiences with a person because planned observations will

    a. cover a wider range of types of behavior.
    b. include a longer time period.
    c. be more economical to obtain.
    d. be more objective and closer to actual behavior.

11. One important virtue of direct observation is that the method

    a. is economical and efficient.
    b. digs into the inner motives of the individual.
    c. can be applied in natural real-life situations.
    d. yields a record of behavior that is directly meaningful.

12. Direct observation would be most useful as a research method if one were trying to answer the question,

    a. How are programs in art staffed in relatively more and less affluent schools?
    b. Is artistic ability related to creativity?
    c. Can an intensive program in art help children to read and write better?
    d. Are there any differences between lower- and middle-class children in their use of artistic media?

13. Direct observation would be most likely to be used in a study of

    a. attitudes of parents toward the school.
    b. teachers' classroom behavior.
    c. children's concepts of life and death.
    d. learning and forgetting under varied conditions of practice.

14. With what sorts of individuals has the method of direct observation proven particularly useful?

    a. Those who are culturally disadvantaged.
    b. Young children.
    c. Individuals from a different culture.
    d. Sophisticated adults.

15. Which of the following is most characteristic of <u>skilled</u> as opposed to <u>unskilled</u> observers in research?

    a. Making generalized descriptions or evaluations rather than noting specific incidents.

b. Interpreting behavior on the basis of limited data rather than waiting for confirmation by subsequent incidents.
c. Recording the child's behavior rather than the personal reaction of the observer.
d. Recording primarily dramatic, unusual, or negative incidents.

16. Sampling of behavior by observational methods can be considered as consisting of two aspects,

a. person sampling and event sampling.
b. person sampling and time sampling.
c. event sampling and time sampling.
d. criterion sampling and person sampling.

17. From the reliability coefficient of a test one can judge

a. how consistently the pupil will maintain his position in the group if an equivalent test is given.
b. how many points the pupil is likely to change if an equivalent test is given.
c. whether the test is measuring what it is supposed to measure.
d. whether the test is related to other significant factors in the individual.

18. "Halo" effect refers to the

a. influence of one rater upon another.
b. tendency to rate a person higher when you know him better.
c. spread of general impression of a person to the rating of many special characteristics.
d. tendency to make ratings too high.

19. Generosity error is illustrated by the fact that

a. few people are ever rated below average.
b. higher ratings are given to close acquaintances.
c. a person who is rated high on one trait is usually rated high on other traits also.
d. one leans over backwards not to be too hard in rating people one doesn't like.

20. Annual achievement tests which are parts of the school testing program, illness records, daily assignments and other routine characteristics and activities of the school are considered to be

a. objective measures.

b. nonreactive measures.
c. reliable measure.
d. standardized measures.

21. As part of a research study a group of 50 teachers were to be rated on the clarity with which they presented concepts of their pupils. Each teacher was to be rated by an observer after a visit of an hour. To make the ratings more reliable, it would probably be most important to

a. extend the visit to two hours.
b. make several different ratings.
c. visit each teacher several times.
d. increase the number of teachers.

22. When subjects perform well merely because they are being observed (and not necessarily because of any effect of treatment), this is considered to be an example of the

a. Hawthorne effect.
b. novelty effect.
c. halo effect.
d. observer effect.

23. In a research study an observer may select categories of events to be observed on the basis of

a. theoretical grounds.
b. practical considerations.
c. empirical evidence.
d. All of the above.

24. For educational purposes, classification schemes are

a. of little use, because they allow for neither prediction nor control of behavior.
b. of limited use, if they are arranged on a meaningful hierarchical basis.
c. potentially very useful, depending on the quality and precision with which categories are defined.
d. extremely useful, given the inadequate development of educational measurement.

25. A questionnaire and an interview are alike in that

a. the situation is standardized and uniform for each respondent.

b. the basic information is supplied by the subject in response to questions.
c. the result is a score or scores arrived at in the same way for each person.
d. the two techniques yield objective information.

26. Separating a group of possible subjects into categories (e.g., males and females) before selecting subjects (by use of a table of random numbers) to receive a particular treatment is known as drawing a

a. stratified random population.
b. simple random sample.
c. unbiased random sample.
d. stratified random sample.

27. It is important to follow up those persons in a survey-sample who have failed to return questionnaires mailed to them? (Assume that the direct-mail method is appropriate.)

a. No, because a certain loss must be expected.
b. No, because such delayed returns would be of doubtful value.
c. Yes, because the sample at hand may be biased.
d. Yes, because the size of sample should be as large as possible.

28. Which of the following is usually the most serious limitation in the use of the mailed questionnaire in descriptive research?

a. Defining a population.
b. Selecting a sample.
c. Obtaining representative responses.
d. Analyzing the data.

29. As compared with laboratory experiments, studies in school settings generally have

a. less control over relevant variables.
b. more control over relevant variables.
c. about equal control over relevant variables.
d. it is impossible to make a comparison in this regard.

30. A procedure in which the researcher observes the child and records his responses is preferably to one in which the child marks an answer sheet in that it

a. is more objective and accurate.
b. is more economical of time and money.

c. permits observation of the process of reaching the answer.
d. corresponds more closely to actual school conditions.

31. In the case of some traits or achievements, we are limited to rather rough subjective evaluations. We should

    a. avoid the evaluation of such traits, because our techniques are subjective.
    b. center our attention on the traits which we can measure objectively.
    c. be content with the subjective procedures, since they are all we have.
    d. use the subjective procedures with caution, while trying to develop improved techniques.

32. One of the serious limitations of the information that is often obtained from systematic behavior observations is that

    a. it usually depends too much on interpretation of what the person is doing.
    b. the specific behaviors may be so isolated and external that it is hard to know what they signify.
    c. it is limited to unreal and artificial situations.
    d. it is dependent upon the cooperation of the subjects.

33. If we are trying to use ratings to provide an evaluation of the effectiveness of teachers, we may replace 3 or 4 broad trait names by a list of 30 or 40 specific behaviors. This may have various results, such as

    I. greater uniformity of meaning from one rater to another.
    II. less relationship of the ratings to actual observations of the ratees' behavior.
    III. more difficulty in using the ratings to remedy individual strengths and weaknesses.

    Which of the above are likely to occur?

    a. I only
    b. I and II only
    c. I and III only
    d. II and III only

34. Which of the following is equivalent to increasing the length of a test?

a. increasing the number of pupils tested.
b. increasing the time allowed for giving the test.
c. increasing the number of raters rating pupils.
d. increasing the range of grades tested.

35. A pupil's interest in entering an experimental art program could probably be determined most adequately by a

a. self-report inventory.
b. situational test.
c. projective method.
d. teacher's rating.

36. Interviews and questionnaires as data collection methods

a. are more effective than observational techniques.
b. reveal only information the subject is willing to report.
c. cannot be considered to have validity.
d. provide no information about past behavior.

37. A research program is being set up to develop tests to use in the selection of students for a new art program. It is probable that the greatest difficulty would arise in

a. selecting promising tests to try out.
b. getting the cooperation of a group of students.
c. working out statistical procedures for determining test validities.
d. obtaining satisfactory measures of success in the program.

38. When a researcher states that a result is significant, he means

a. the effect is a practically important one.
b. the scores are highly correlated.
c. the result is unlikely to be a chance occurrence.
d. the sample values are different from one another.

39. The statistical methods to be used in a study should be considered

a. before the hypotheses have been formulated.
b. as the study is being designed.
c. after the data have been collected.
d. after the data have been tabulated.

40. When a researcher prepares a well-designed set of charts showing the results of a study, the educational significance of the results should be

a. immediately apparent from the charts.
b. supplied in accompanying captions or text.
c. left to the reader to work out for himself.
d. a matter for each individual's own education philosophy.

1970 NAEA

Preconference Educational Research Training Program
In Art Education

Key

1. d
2. c
3. b
4. d
5. d
6. a
7. c
8. a
9. a
10. d
11. c
12. d
13. b
14. b
15. c
16. c
17. a
18. c
19. a
20. b
21. c
22. a
23. d
24. c
25. b
26. d
27. c
28. c
29. a
30. c
31. d
32. b
33. a
34. c
35. a
36. b
37. d
38. c
39. b
40. b

APPENDIX E

PARTICIPANTS EVALUATION FORM

1970 NAEA

Preconference Educational Research Training Program In Art Education

Please complete this form and return it to the Evaluator, William Rabinowitz, at the end of the program.

. . . . . .

1. How would you evaluate the organization of the program?

_____ Excellent organization in clear and meaningful sequence

_____ Generally well-organized

_____ Adequately organized, but could have been better

_____ Inadequately organized, too much confusion

_____ Very disorganized, no apparent structure

Comments: _____

_____

_____

2. How realistic and attainable were the objectives of the program?

_____ Very realistic and easily attainable

_____ Most of the objectives were clearly attainable

_____ Many of the objectives were attainable, but some were not

_____ Most of the objectives were not attainable

_____ The objectives of the program were clearly unrealistic

Comments: _____

_____

_____

3. To what extent did the program meet your prior expectations?

_____ The program exceeded my expectations

_____ My expectations were generally well-met

_____ The program was O.K., but not all my expectations were met

_____ I expected more from the program than I obtained

_____ The program completely failed to meet my expectations

Comments: _____

_____

_____

4. To what extent did the program deal with questions and problems that are applicable to your needs?

_____ The program was exceptionally well-related to my needs

_____ The program was generally well-related to my needs

_____ The program was adequate in terms of my needs

_____ The program was only slightly related to my needs

_____ The program was completely unrelated to my needs

Comments: _____

_____

_____

5. Was the length of the program appropriate?

_____ Considering what was covered, the program was too long

_____ The length of the program was just right

_____ Considering what was covered, the program was too short

Comments: _____

_____

_____

6. Was the difficulty level of the program appropriate?

_____ Considering my background, the program was too difficult

_____ The difficulty level of the program was just right

_____ Considering my background, the program was too easy

Comments: _____

_____

_____

7. In general, how would you evaluate the quality of the instruction?

_____ Outstanding

_____ Good

_____ Satisfactory

_____ Inadequate

_____ Poor

Comments: _____

_____

_____

8. Please indicate in the space below one or two ways in which you plan to use ideas or skills acquired during the program.

_____

_____

_____

_____

_____

_____

9. Would you recommend that the NAEA continue to sponsor research programs in the future?

_____ Yes _____ No

10. Please indicate in the space below your suggestions for topics or problem areas for future programs. You may also indicate any other general suggestions for conducting future programs.

_____

_____

_____

_____

_____

_____

_____

Signature: _____ (Note: This is optional.)